



Cognitive processing of nonverbal cues in earnings conference calls: The association between CEO vocal emotion and stock market reactions

Baojun Ma^a, Lingyun Zhou^b, Yao Mu^{c,*}, Yi Chen^c, Jian Zhang^c

^a Business School, Hohai University, Nanjing 211100, China

^b Shanxi Finance and Taxation College, Taiyuan 030024, China

^c Key Laboratory of Brain-Machine Intelligence for Information Behavior (Ministry of Education and Shanghai), School of Business and Management, Shanghai International Studies University, Shanghai 201620, China

ARTICLE INFO

Keywords:

CEO vocal emotion

Stock market reactions

Two-dimensional emotion space

Speech emotion recognition

ABSTRACT

As important nonverbal cues, CEO vocal emotion in earnings conference calls conveys incremental information about firms. This study investigates the association between CEO vocal emotion and stock market reactions from the perspective of investor cognitive processing. Particularly, speech emotion recognition techniques are applied to measure CEO vocal valence and arousal in the two-dimensional emotion space and determine the dynamic emotion reversals. Based on a large panel dataset, we conduct various empirical tests and robustness checks to find that CEO vocal valence in earnings conference calls is positively associated with corporate cumulative abnormal returns over the next two trading days. CEO vocal arousal nonlinearly moderates this association, which is strongest at medium levels of arousal. Furthermore, a high frequency of emotion reversals strengthens the main association. This study provides in-depth insights into the role of CEO vocal emotion in market, and the validation of boundary conditions sheds light on mechanism exploration.

1. Introduction

As a primary form of voluntary disclosure, earnings conference calls play a crucial role in reducing information asymmetry between firms and investors (Bergh et al., 2019). Given that indicators, reports, and news regarding corporate performance are typically disseminated through other routine channels, the distinctive and significant value of these voluntary calls lies in their live and interactive nature, which allows market participants to gather additional cues from the audio information (Call et al., 2024). Enlightened by signaling theory, earnings conference calls can be interpreted as one-way communication, in which chief executive officers (CEOs), as signallers, hold private information about firms, while investors or analysts, as receivers, lack this information and thus rely on the transmitted cues by CEOs to guide their decisions (Moradi et al., 2023; Ning et al., 2024). Prior research finds that verbal content in these calls can influence stock market reactions, even to a greater extent than common financial metrics (Myšková & Hájek, 2020). With CEOs engaging investors through various media channels, recent studies emphasize the importance of nonverbal signals or cues that cannot be captured in texts, such as CEO vocal emotion, and

demonstrate that such cues may provide investors with deeper insights into a firm's condition (Doey & de Jong, 2025; Elliott et al., 2024).

During earnings conference calls, particularly within the management discussion and analysis (MD&A) section, CEOs convey firm information on past performance, current status, and future prospects with certain vocal emotion (Chakraborty & Bhattacharjee, 2020). Investors often treat these nonverbal cues as incremental messages, which provide complementary insights that go beyond the disclosed content (Elliott et al., 2024). By interpreting vocal emotion, they infer CEOs' implicit assessments of firm performance and prospects (Ning et al., 2024; Zhang et al., 2024); that is, positive emotion may signal a favorable internal outlook, while negative emotion can imply underlying concerns. Therefore, CEO vocal emotion shapes investors' perception and cognition of firm conditions, ultimately influencing their investment decision-making (Chou et al., 2018; Mayew & Venkatachalam, 2013). For instance, Mayew and Venkatachalam (2012) find that investor responses to CEO vocal emotion during earnings conference calls are significantly associated with stock market reactions, because vocal excitement serves as a proxy for positive emotion and signals good performance, whereas cognitive dissonance, as a proxy for negative emotion, suggests poor

* Corresponding author.

E-mail addresses: mabaojun2003@163.com (B. Ma), zhoulingyun@sxft.edu.cn (L. Zhou), muy@shisu.edu.cn (Y. Mu), chenyied@outlook.com (Y. Chen), jianzhang@shisu.edu.cn (J. Zhang).

<https://doi.org/10.1016/j.jbusres.2026.116569>

Received 19 July 2025; Received in revised form 8 September 2026; Accepted 25 September 2026

Available online 29 September 2026

0148-2963/© 2026 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

prospects. Moreover, Price et al. (2017) confirm that investors can quickly react to nonverbal cues revealed by CEO vocal emotion, which are strongly correlated with the firm's stock price movements measured by cumulative abnormal returns (CAR).

However, previous research often treats CEO vocal emotion as a general construct and has rarely considered the underlying rationales through which it relates to stock market reactions. This hinders in-depth investigations into its role in the stock market and the exploration of potential mechanisms. Therefore, this study employs a novel perspective of investor cognitive processing to identify and address the limitations of existing research. First, prior studies predominantly focus on the valence of CEO vocal emotion, that is, the positive or negative polarity (Mayew & Venkatachalam, 2012; Price et al., 2017). Nevertheless, during earnings conference calls, CEO vocal valence is not conveyed in isolation but interacts with other emotional characteristics, which jointly shape the outcomes of investors' cognitive processing (Chung & Lee, 2019). Hence, it is necessary to consider a richer set of vocal emotional dimensions and the complex interplay between them. Second, most existing studies measure CEO vocal emotion as an aggregate over the entire call, with little attention paid to emotional fluctuations (Mayew & Venkatachalam, 2013; Price et al., 2017). For investors, cognitive processing is not a static state; rather, they dynamically reallocate their cognitive resources in response to changes in the CEO's vocal emotion (Nabi & Green, 2015). Therefore, introducing fine-grained fluctuations in the dynamic emotional flow can help explain stock market reactions more effectively. Third, much of the previous research aims either at predicting market reactions by extracting CEO vocal characteristics or exploring their main effects on the stock market (Ewertz et al., 2025; Yang et al., 2023). To gain deeper theoretical insights, it is essential to further investigate the underlying mechanisms and boundary conditions, such as analyzing how and when CEO vocal emotion influences stock market reactions by shaping investors' cognitive processing.

Grounded on the limited capacity model of motivated mediated message processing (LC4MP) (Lang, 2000; 2006), this study investigates the association between CEO vocal emotion in earnings conference calls and stock market reactions. This theoretical perspective has been widely applied to examine investor responses to corporate disclosures (Chung & Lee, 2019; 2020; DiRusso & Myrick, 2021), which facilitates to explain how CEO vocal emotion may activate investors' motivational systems and shape their cognitive processing. Specifically, to comprehensively capture the emotional nuance, we utilize the well-established two-dimensional (2D) emotion space model (Russell, 2003; Sykora et al., 2022) to characterize CEO vocal emotion by two continuous metrics, namely valence and arousal. Valence reflects the emotional polarity of voice, while arousal captures its intensity or degree of activation (Chung et al., 2015; Tanghe et al., 2010). Additionally, from a dynamic viewpoint, we incorporate emotion reversals to assess the CEO's vocal valence shifts across sentences during the call. Using a pretrained speech emotion recognition (SER) model to achieve automatic and rapid quantification (Wagner et al., 2023), this study explores the association between CEO vocal valence and stock market reactions, which is measured by CAR over the next two trading days after the call (CAR (0,1)), and the moderating effects of CEO vocal arousal and the frequency of emotion reversals. Empirical tests are conducted with two-way fixed effects models and difference-in-differences (DiD) analysis on a large dataset spanning the period of 2017–2022 and containing about 1,000 U.S.-listed firms and more than 1,300 CEOs. To identify the role of CEO vocal emotion effectively, the tests control for a wide range of verbal and acoustic features, as well as characteristics of the calls, firms, and CEOs, which extensively capture the potential factors related to stock market reactions. The results show that CEO vocal valence in earnings conference calls is positively associated with the firm's CAR (0,1). CEO vocal arousal nonlinearly moderates this association, which is strongest at medium levels of arousal. Furthermore, a high frequency of emotion reversals strengthens the association between CEO vocal valence and stock market reactions. These findings remain consistent in

multi-view robustness checks of alternative settings such as recording sections, key variable measures, additional controls, and fixed effects.

The main contributions of this work are threefold. First, unlike previous studies that focus mainly on CEO vocal valence, this study employs SER techniques to measure emotion in a 2D continuous space, so as to investigate the complex, nonlinear interaction between CEO vocal arousal and valence and its association with stock market reactions. Hence, this work expands existing findings with the enriched dimensions of vocal emotion. Second, in contrast to earlier work that typically assesses the CEO's aggregated vocal emotion over the entire call, this study incorporates the moderating effect of emotion reversals within a dynamic emotional flow, which supplements fine-grained insights into the role of CEO vocal valence. Third, this work broadens the scope of existing research by exploring the potential mechanism of the association between CEO vocal valence and stock market reactions, that is, CEO vocal valence acts as a nonverbal cue to activate investors' motivational systems, thereby altering their cognitive processing and subsequent investment decisions. The empirical validation of the moderating effects of CEO vocal arousal and emotion reversal frequency not only implies previously unexplored boundary conditions, but also provides evidence for our mechanism exploration.

2. Literature review

To reduce information asymmetry with investors, corporations typically make public disclosures by textual documentation, such as annual report, IPO prospectus, and announcement of major issues. Based on the content and emotion conveyed in texts, investors can better understand the firm conditions (Ahmed et al., 2024; Uparna & Bingham, 2022).

Unlike text-based corporate disclosures, the voluntary earnings conference calls enrich the information available to market participants with audio recordings, allowing for richer analysis of firm performance (Call et al., 2024; Chou et al., 2018). Numerous empirical studies focus on the sentiment of verbal content during the calls and find that it can be a more significant predictor of stock market reactions than common financial measures (Amoozgar et al., 2020; Mysková & Hájek, 2020). Considering the incremental and distinctive messages delivered by voice, recent work pays increasing attention to the role of nonverbal cues such as CEO vocal emotion (Doey & de Jong, 2025; Elliott et al., 2024). As vocal emotion signals the CEO's implicit interpretation of firm conditions, investors' perception and cognition of the firm's future performance can be affected, leading to a subsequent influence on stock market reactions (Ning et al., 2024; Zhang et al., 2024).

Regarding the role of CEO vocal emotion in earnings conference calls, there are mainly two streams of prior studies. The first stream extracts acoustic features, such as spectral features and frequency components, from audio recordings to construct proxies of vocal emotion. For instance, Mayew and Venkatachalam (2012) extract CEO vocal excitement as a proxy of positive emotion and cognitive dissonance as negative signals, and find that both proxy variables are significantly associated with investor reactions reflected in stock price movements. Price et al. (2017) adopt the layered voice analysis (LVA) software to quantify excitement and the presence of local low frequencies in CEO voices so as to construct emotion variables, which show high correlations with stock market reactions measured by CAR after the call. However, these acoustic features are commonly mapped to a single valence dimension of vocal emotion, which hardly supports the nuanced and comprehensive exploration on emotional associations. As software like LVA typically operates as a black box, the validity and reliability of such quantification is also uncertain, especially for theoretical investigation to infer the role of vocal emotion on market responses (Eriksson & Lacerda, 2007; Lacerda, 2009). In addition, the consistency and robustness of emotion measurement based on acoustic features has undergone limited empirical scrutiny (Mayew & Venkatachalam, 2013; Schuller et al., 2011).

The second stream applies machine learning models that are trained on labeled speech corpora to classify recordings into discrete emotional states, such as joy, anger, and sadness. For example, Hajek (2022) uses a SER model based on convolutional neural network (CNN) to obtain discrete emotional states of CEOs for financial distress prediction. Similarly, Ewertz et al. (2025) adopt a CNN-based model to classify CEO vocal emotion during earnings conference calls and show its superior prediction power for firm performance and analyst reactions. To further improve the measurement, Yang et al. (2023) develop a vocal emotion classification system with a two-stage long short-term memory network for risk prediction. Nevertheless, these discrete states have limitations in characterizing the complexity and richness of human vocal emotion. Although distinct states may manifest differences on multiple emotional dimensions (e.g., valence and arousal), the discrete characterization can hardly capture fine-grained associations between CEO vocal emotion and stock market reactions. In particular, given that the interactions among emotional dimensions may be nonlinear (Chung & Lee, 2019; Chung & Sparks, 2016), only continuous measurement can uncover such complex relationships. Prior studies have also found that incorporating discrete emotional states as a supplement to acoustic features fails to enhance or even decreases the accuracy of prediction methods (Hajek, 2022; Yang et al., 2023).

In summary, existing research has yet to adequately account for the complex interplay between different dimensions of CEO vocal emotion, nor has it measured fine-grained emotional fluctuations from a dynamic perspective. Moreover, the underlying mechanisms and boundary conditions through which vocal emotion relates to stock market reactions remain insufficiently explored. These insights motivate our study to delve into the association between CEO vocal emotion in earnings conference calls and stock market reactions from the theoretical lens of investor cognitive processing, so as to yield richer and more robust research findings.

3. Hypotheses

3.1. CEO vocal valence and stock market reactions

During earnings conference calls, mainly in the MD&A section, CEOs communicate firm information to investors and exhibit varying vocal emotion, from which investors can infer CEOs' implicit interpretation of firm performance in the past and future (Moradi et al., 2023; Ning et al., 2024). As important cues in addition to verbal content, CEO vocal emotion may intentionally or unintentionally disclose private information not reflected in the textual transcripts, because vocal emotion is generally less controllable and more prone to leakage (Dietrich et al., 2019; Schoofs & Claeys, 2021).

This study adopts the 2D emotion space, the validity and reliability of which has been verified and widely accepted (Tanghe et al., 2010), to measure fine-grained CEO vocal emotion in earnings conference calls. Specifically, the continuous dimension of valence reflects the emotional polarity of voice, and a higher valence implies a more positive versus negative emotion (Russell, 2003; Sykora et al., 2022). As indicated by LC4MP that is grounded in motivated cognition theory, emotion and cognition are inherently intertwined, enabling people to assess and react to emotion-related signals through two core motivational systems, namely appetitive and aversive (Khandeparkar et al., 2021; Lang, 2000; 2006). Therefore, when investors are exposed to the CEO voice with positive valence, their appetitive system can be activated, which shapes their cognition of the firm's positive prospects and thus encourages favorable market reactions (Wang & Lang, 2012). In contrast, the negative vocal valence of CEO can trigger the aversive system of investors, leading to their negative interpretation of the firm's future performance and cautious investment decisions (Lang, 2000). Consistent with this route, previous studies based on emotion proxies have found

that the positive vocal emotion of CEO is associated with enhanced investment intentions and short-term stock market reactions, while the negative vocal emotion is related to muted investor responses (Mayew & Venkatachalam, 2012; 2013; Price et al., 2017).

Therefore, we propose the following hypothesis:

Hypothesis 1 CEO vocal valence in earnings conference calls is positively associated with stock market reactions.

3.2. Moderating effect of CEO vocal arousal

As prior work argues that emotion with similar valence can result in different effects on receiver decision-making, the importance of finer-grained distinctions of emotion becomes prominent (Tiedens & Linton, 2001), especially considering the other dimension (i.e., arousal) in the 2D emotion space (Chung et al., 2015; Robinson et al., 2004). For example, Rhudy et al. (2008) conduct an emotional picture-viewing experiment, which demonstrates that arousal effectively moderates the association between emotional valence and subsequent responses.

For earnings conference calls, the interaction between CEO vocal valence and arousal can be interpreted from the perspective of investor cognitive processing. According to LC4MP, CEO vocal emotion shapes investor cognition by activating their motivational systems (Lang, 2000; 2006). Although valence determines which system is activated (i.e., appetitive or aversive), emotional arousal modulates the intensity of motivational activation (Chung & Lee, 2019; 2020). Furthermore, related studies have shown that arousal affects the allocation of cognitive resources in a nonlinear way (Lang, 2006; 2009). Specifically, for the appetitive system, people's cognitive resource allocation generally increases with arousal, but the effect flattens as arousal grows extreme. For the aversive system, negative emotion triggers greater cognitive processing at low to medium levels of arousal, and the influence declines at high arousal (Chung & Lee, 2019; Chung & Sparks, 2016). Therefore, for CEO vocal emotion during earnings conference calls: on one hand, low arousal may reduce investor attention, which suppresses the activation of motivational systems, limits investors' cognitive processing, and thus weakens the effect of CEO vocal valence (O'hanlon, 1981); on the other hand, excessive arousal not only induces stress and impairs cognitive functions but also signals irrationality of the CEO, leading to an impression of exaggeration or inauthenticity, which may trigger investor suspicions and constrain the potential influence of CEO vocal valence on market responses (Yin et al., 2017).

In summary, we argue that the increase in CEO vocal arousal enhances investor cognitive processing until it reaches an optimal level to furthest amplify the association between vocal valence and stock market reactions, and this effect diminishes as arousal continues to increase. Thus, we propose:

Hypothesis 2 CEO vocal arousal nonlinearly moderates the association between CEO vocal valence and stock market reactions, and the positive association is strongest at medium levels of arousal.

3.3 Moderating effect of the frequency of emotion reversals

The emotion dynamics theory posits that people's emotional expressions continuously fluctuate over time as a flow, so aggregated measures cannot fully capture the complexity and nuances of emotion (Jiang et al., 2019; Kuppens & Verduyn, 2017). During earnings conference calls, CEO vocal emotion also fluctuates in arousal and may even reverse the valence, which leads to varying effects on investor cognition and thus requires deeper insights into the dynamic emotional flow (Aurier & Guintcheva, 2015; Keene & Lang, 2016).

This study considers the significant polarity shifts in CEO vocal valence, namely emotion reversals, during earnings conference calls to depict the dynamic nature of emotional flow (Yiu et al., 2024; Zhang et al., 2025). Previous work has found that emotion reversals increase

audience engagement in media exposures compared to a single-polarity scenario, which significantly enhances persuasion and contributes to the exposure effect (Alam & So, 2020; Guido et al., 2018; Knight et al., 2024). By interpreting LC4MP from a dynamic viewpoint, CEO emotion reversals, given their across-polarity nature, can sequentially activate different motivational systems of investors (e.g., shifting from the appetitive to the aversive system or vice versa). This triggers oriented responses and a dynamic reallocation of cognitive resources during the call (Nabi & Green, 2015; Wang & Lang, 2012). By sustaining investor attention and cognitive processing at a heightened level, frequent CEO emotion reversals can increase investors' engagement in earnings conference calls and thus enhance the transmission of emotional cues, which is ultimately reflected in stock market reactions (Keene & Lang, 2016; Schmidt et al., 2023).

In summary, a high frequency of emotion reversals leads to increased cognitive processing and engagement of investors, potentially amplifying the influence of CEO vocal valence on their cognition and investment decisions. So, we propose:

Hypothesis 3 A high frequency of emotion reversals strengthens the positive association between CEO vocal valence and stock market reactions.

Methodology

4.1. Sample

To investigate the association between CEO vocal emotion and stock market reactions, we collected available quarterly earnings conference calls of U.S.-listed firms from Seeking Alpha¹ (i.e., one of the largest investor-oriented platforms in the U.S.) in May 2023. After preliminary screening, we obtained 52,010 calls between 2017 and 2022 with the CEO in attendance, including the audio recording and transcript for each. Next, based on company name, stock code, and CEO name, these calls were matched with firm-level financial data from the Quarterly Compustat database, historical stock prices from the CRSP database, and CEO demographic characteristics from the Execucomp database. Given the pool of 19,599 matched observations, we filtered for firms within the S&P 1500 index, yielding a refined sample of 16,901 observations. Then, to ensure data quality and comparability, we excluded financial firms (SIC 6000–6999), utilities (SIC 4900–4949), and observations with missing or invalid values. The final panel dataset consists of 10,787 firm–quarter–CEO records from 998 firms and 1,338 CEOs.

4.2. Measures

4.2.1. Stock market reactions

In this study, we measure the dependent variable (i.e., stock market reactions) with the firm's CAR over the next two trading days after the earnings conference call, which is denoted as CAR(0,1). For previous studies on CEO vocal emotion, CAR is widely used to capture investor responses (Mayew & Venkatachalam, 2012; Price et al., 2017), and CAR (0,1) is especially effective in reflecting the short-term stock market reactions to public disclosures by firms (de Amicis et al., 2021; Pantzalis & Ucar, 2014). In alignment with the common practice in prior literature, calls held after the general stock market closing time (16:00 EST) are assigned to the next trading day (Berkman & Truong, 2009; Liu et al., 2023). As market reactions to a call would manifest at the earliest on the following trading day, such adjustments for the lagged effects help to diminish bias into statistical analysis on stock market reactions. Furthermore, the return estimation is performed using a 120-day time window starting from 130 days before the call (Sprenger et al., 2014). Specifically, the CAR(0,1) variable is calculated with the S&P 1500 index by Equation (1).

$$CAR(0,1) = \sum_{d=0}^1 R_{i,d} - E(R_{i,d}), \quad (1)$$

where $R_{i,d}$ denotes the actual stock return for firm i on day d , and $E(R_{i,d})$ is the corresponding estimated return.

4.2.2. CEO vocal emotion

To measure CEO vocal emotion in the 2D continuous space, this study adopts a pretrained SER model (Wagner et al., 2023) to quantify CEO vocal valence (i.e., the independent variable) and arousal (i.e., the moderator variable) automatically and rapidly. On one hand, the deep learning-based measurement exhibits higher scalability and time efficiency than human rating (Siegert et al., 2014), which is ideal for empirical analyses that require a large sample size to derive robust insights. On the other hand, the adopted model has been validated to outperform other state-of-the-art methods in accuracy, robustness, and fairness of vocal emotion quantification (Wagner et al., 2023), and widely used in real-world applications such as automatic assessment of attention-deficit or hyperactivity disorder (Spiesberger et al., 2024) and suicide risk (Amiriparian et al., 2024).

Specifically, given each earnings conference call in the dataset, the automated workflow to measure CEO vocal emotion consists of six steps as shown in Fig. 1. (1) Tokenize the earnings conference call transcript into sentences using a public Python library named Natural Language Toolkit²; (2) Perform forced alignment with Aeneas³ and eSpeak⁴ (i.e., a popular text-to-speech software) to align CEO audio recording with the textual transcript; (3) Apply regular expression matching to locate CEO narratives in the MD&A section and obtain the corresponding time-stamps; (4) Segment the audio recording based on timestamps using the Pydub⁵ library to isolate vocal sentences; (5) Predict vocal valence and arousal scores at the sentence level with the adopted SER model (Wagner et al., 2023); (6) Calculate the average valence and arousal scores of all sentences in the MD&A section and then perform z-score normalization on both metrics. After scaling the normalized values to the range of [0,1], the overall CEO vocal valence (AUDIOV) and arousal (AUDIOA) are derived.

4.2.3. Frequency of emotion reversals

As stated, this study considers the CEO's dynamic emotional flow in addition to the overall valence. Enlightened by prior literature (Berger et al., 2021; Yiu et al., 2024; Zhang et al., 2025), we define emotion reversal as a significant polarity shift in CEO vocal valence across sentences. Specifically, to adapt to the data and domain context, valence values above the sample median can be regarded positive, while those below are rather negative. In two adjacent sentences, if CEO vocal valence shifts from positive to negative (or contrary) and the valence difference between the two sentences exceeds two standard deviations (i.e., 0.2 in our dataset), then the shift accounts for an emotion reversal. Therefore, the frequency of emotion reversals (i.e., the moderator variable, REVERSAL) is calculated as the ratio of the times of CEO emotion reversals to the total number of sentences in the MD&A section.

4.2.4. Control variables

To account for factors that may be associated with CAR(0,1) after earnings conference calls, we incorporate a wide range of control variables. For calls, we control for verbal content emotion in the MD&A section by assessing the sentence-level valence and arousal scores in transcript texts based on the widely used VADER model (Hutto & Gilbert, 2014) and NRC-VAD lexicon (Mohammad, 2018), respectively, and calculating the average textual valence (TEXTV) and arousal

¹ <https://www.seekingalpha.com/earnings/earnings-call-transcripts>

² <https://www.nltk.org>

³ <https://github.com/readbeyond/aeneas>

⁴ <https://espeak.sourceforge.net>

⁵ <https://pypi.org/project/pydub>

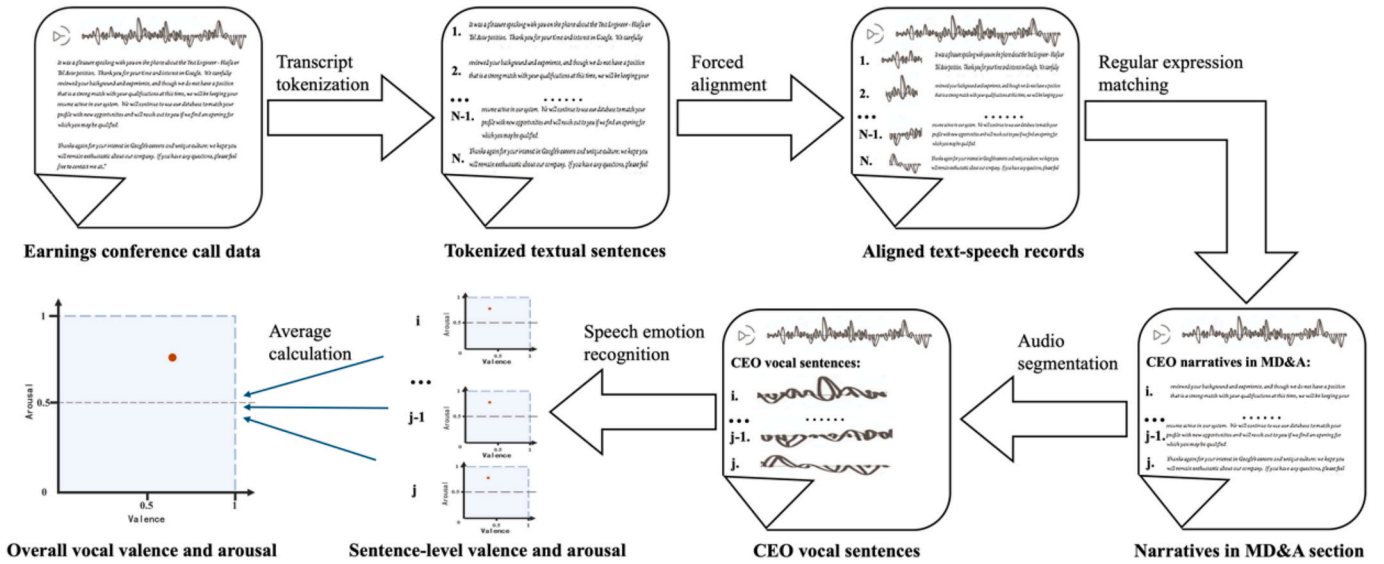


Fig. 1. Automated workflow to measure CEO vocal valence and arousal.

(TEXTA). Enlightened by previous studies on CEO vocal emotion, we also control for five acoustic features extracted by the Praat software (Boersma & Van Heuven, 2001), including the mean and standard deviation of fundamental frequency ($F_0(\text{Mean})$, $F_0(\text{Std})$), the relative average vocal perturbation (JITTER), moment-to-moment amplitude variation (SHIMMER), and harmonic-to-noise ratio (HNR) of CEO speech in the MD&A section (Mayew & Venkatachalam, 2012; Throckmorton et al., 2015). In addition, we control for basic characteristics following the related literature (Pollock et al., 2024), including the total duration of the MD&A section (TIME), the proportion of the CEO's speaking time (PROPORTION), the number of executives in attendance (EXECUTIVES), and the number of analysts (ANALYSTS).

For firms, we control for financial performance metrics such as return on assets (ROA), market value (MCAP), sales growth (GROWTH), book-to-market ratio (BTM), and firm leverage (LEV). Following Price et al. (2012), we incorporate controls for other factors that may influence investor behavior, including market uncertainty (VOLATILITY), unexpected earnings (SURP), and share trading volume (VOLUME). Moreover, for CEOs, we control for their gender (GENDER), age (AGE), and years of tenure (TENURE).

To mitigate the effect of outliers, we apply winsorization at the 1st and 99th percentiles to all continuous variables, following the common treatment in prior studies (Leone et al., 2019). The definition and calculation of variables is summarized in Table 1.

4.3. Models

Based on the panel dataset, this study uses a two-way fixed effects model to estimate the association between CEO vocal valence in earnings conference calls and stock market reactions. As shown in Equation (2), to mitigate endogeneity concerns, firm fixed effects are employed to address the unobserved heterogeneity of corporations, and year–quarter fixed effects are included to control for macroeconomic variations both across years and across quarters. In addition, we apply firm-clustered standard errors to address potential cross-sectional correlations and time-series autocorrelations so as to enhance the robustness of estimation.

$$CAR(0,1)_{i,t} = \beta_0 + \beta_1 AUDIOV_{i,t} + \beta_2 AUDIOA_{i,t} + \beta_3 REVERSAL_{i,t} + \beta_4 AUDIOA_{i,t}^2 + \alpha X_{i,t} + \gamma_i + \delta_t + \varepsilon_{i,t}, \quad (2)$$

where i denotes the firm, t denotes the year–quarter, γ_i and δ_t are the firm and year–quarter fixed effects, respectively, and $\varepsilon_{i,t}$ is the error. $X_{i,t}$

represents the control variables, α represents their coefficients to be estimated, and $\alpha X_{i,t} = \sum_{k=1}^K \alpha_k X_{k,i,t}$ denotes a linear combination, in which $K = 22$ is the total number of control variables. As discussed in Hypothesis 1, the coefficient β_1 is hypothesized to be positive, implying the positive association between CEO vocal valence (AUDIOV) and stock market reactions ($CAR(0,1)$).

To evaluate the moderating effects of CEO vocal arousal and the frequency of emotion reversals on the association between CEO vocal valence and stock market reactions, we construct additional regression models for estimation, as shown in Equations (3) and (4), respectively.

$$CAR(0,1)_{i,t} = \beta_0 + \beta_1 AUDIOV_{i,t} + \beta_2 AUDIOA_{i,t} + \beta_3 REVERSAL_{i,t} + \beta_4 AUDIOA_{i,t}^2 + \beta_5 AUDIOV_{i,t} \times AUDIOA_{i,t}^2 + \beta_6 AUDIOV_{i,t} \times AUDIOA_{i,t} + \alpha X_{i,t} + \gamma_i + \delta_t + \varepsilon_{i,t}, \quad (3)$$

$$CAR(0,1)_{i,t} = \beta_0 + \beta_1 AUDIOV_{i,t} + \beta_2 AUDIOA_{i,t} + \beta_3 REVERSAL_{i,t} + \beta_4 AUDIOA_{i,t}^2 + \beta_7 AUDIOV_{i,t} \times REVERSAL_{i,t} + \alpha X_{i,t} + \gamma_i + \delta_t + \varepsilon_{i,t}. \quad (4)$$

As illustrated in Hypothesis 2, we focus on the coefficient β_5 of the squared interaction term $AUDIOV_{i,t} \times AUDIOA_{i,t}^2$ in Equation (3), which is hypothesized to be negative to suggest that CEO vocal arousal non-linearly moderates the association between AUDIOV and $CAR(0,1)$, and the association is strongest at medium levels of AUDIOA. To test Hypothesis 3, we examine the coefficient β_7 of the interaction term $AUDIOV_{i,t} \times REVERSAL_{i,t}$ in Equation (4), which is hypothesized to be positive, implying that a high frequency of emotion reversals strengthens the association between CEO vocal valence and stock market reactions.

5. Results

5.1. Descriptive statistics and correlation analysis

Table 2 presents the descriptive statistics of variables. As shown in the table, the mean and standard deviation of $CAR(0,1)$ are 0 and 0.07, respectively. The mean and standard deviation of AUDIOV are 0.58 and 0.10, respectively. Regarding the moderator variables, the mean values of AUDIOA and REVERSAL are 0.49 and 0.24, respectively.

The bivariate correlations of variables are also calculated. Specifically, by two-tailed statistical tests, CEO vocal valence (AUDIOV) and stock market reactions ($CAR(0,1)$) are significantly and positively

Table 1
Definition and calculation of variables.

Variable	Definition and calculation
CAR(0,1)	Cumulative abnormal returns over the next two trading days after the earnings conference call, calculated with the S&P 1500 index.
AUDIOV	CEO vocal valence, calculated as the average valence score in the MD&A section assessed by the SER model, followed by z-score normalization and scaling to the range of [0,1].
AUDIOA	CEO vocal arousal, calculated as the average arousal score in the MD&A section assessed by the SER model, followed by z-score normalization and scaling to the range of [0,1].
REVERSAL	Frequency of emotion reversals, calculated as the ratio of the times of significant shifts in CEO vocal valence to the total number of sentences in the MD&A section.
TEXTV	Textual valence, calculated as the average of normalized valence scores of all transcript sentences in the MD&A section that are assessed by the VADER model.
TEXTA	Textual arousal, calculated as the average of arousal scores of all transcript sentences in the MD&A section that are assessed with the NRC-VAD lexicon.
F ₀ (Mean)	Mean of fundamental frequency, calculated as the natural logarithm of average fundamental frequency of all sentences in the MD&A section assessed by the Praat software.
F ₀ (Std)	Standard deviation of fundamental frequency, calculated as the natural logarithm of standard deviation of fundamental frequency in the MD&A section assessed by the Praat software.
JITTER	Relative average vocal perturbation, calculated as the average of JITTER scores of all sentences in the MD&A section assessed by the Praat software.
SHIMMER	Moment-to-moment amplitude variation, calculated as the average of SHIMMER scores of all sentences in the MD&A section assessed by the Praat software.
HNR	Harmonic-to-noise ratio, calculated as the natural logarithm of average HNR score of all sentences in the MD&A section assessed by the Praat software.
TIME	Time duration, calculated as the natural logarithm of the total duration (in seconds) of CEO speech in the MD&A section.
PROPORTION	Proportion of CEO's speaking time, calculated as the ratio of the CEO's speaking time to the total duration of the MD&A section.
EXECUTIVES	Number of executives, calculated as the natural logarithm of the number of executives participating in the earnings conference call.
ANALYSTS	Number of analysts, calculated as the natural logarithm of the number of analysts participating in the earnings conference call.
ROA	Firm's return on assets, calculated as the ratio of quarterly operating income after depreciation to total assets.
MCAP	Firm's market value of equity, calculated as the natural logarithm of the product (in millions) of quarterly stock price and the number of shares outstanding.
GROWTH	Firm's sales growth, calculated as the percentage increase in total sales compared to the same quarter in the last year.
BTM	Firm's book-to-market ratio, calculated as the ratio of quarterly stockholders' equity to the market value of equity (i.e., MCAP).
LEV	Firm leverage, calculated as the ratio of quarterly total debt to the total market assets value, which is the sum of total debt and the market value of equity.
VOLATILITY	Market uncertainty, calculated as the standard deviation of daily returns in a 90-day time window ending at 10 days before the earnings conference call.
SURP	Unexpected earnings, calculated as the ratio of quarterly earnings-per-share minus that of the same quarter in the last year to stock price at the end of the lagged quarter.
VOLUME	Share trading volume, calculated as the total share trading volume (in millions) on the day of the earnings conference call.
GENDER	CEO's gender, which takes the value of one for female, and zero for male.
AGE	CEO's age, calculated as the natural logarithm of age.
TENURE	CEO's tenure, calculated as the natural logarithm of the time duration (in years) that the executive has served as the corporate CEO.

correlated ($r=0.033$, $p<0.01$). AUDIOV exhibits significant positive correlations with AUDIOA ($r=0.146$, $p<0.01$) and REVERSAL ($r=0.100$, $p<0.01$). Furthermore, we calculate the variance inflation factors (VIFs) to assess multicollinearity. For the baseline model without interaction terms, the highest VIF value is 2.68, and the average is 1.58, which is well below the threshold of 10, indicating no serious multicollinearity

concerns (Cohen et al., 2003). For the full model, to decrease the VIFs of interaction terms, the corresponding variables (AUDIOV, AUDIOA, and REVERSAL) are centered for the regression referring to widely-accepted guides (Hayes, 2013), and the highest derived VIF is 1.65 with an average of 1.58.

5.2. Hypothesis tests

For hypothesis tests, this study estimates five two-way fixed effects regression models. The results are summarized in Table 3, in which each column refers to a distinct model specification. The full regression output including point estimates and standard errors for all control variables is presented in Table A1 of the Appendix.

In the Model 0 column of Table 3, only the control variables are included in the regression. The coefficients reveal that the moderator variables (AUDIOA and REVERSAL) have statistically insignificant associations with the dependent variable (CAR(0,1)).

In Model 1, the independent variable (AUDIOV) is added to the regression to test Hypothesis 1, as specified in Equation (2). The results demonstrate that CEO vocal valence in earnings conference calls is positively and significantly associated with stock market reactions ($\beta=0.031$, $p<0.01$), supporting Hypothesis 1. Suppose that other variables take their fixed mean values, an increase in AUDIOV from one standard deviation below the mean to one standard deviation above the mean can lead to a 93.94% increase (from 0.0066 to 0.0128) in the predicted CAR(0,1), which is remarkable.

In Model 2, we add the interaction terms with regard to AUDIOA to the regression as shown in Equation (3), and find that CEO vocal arousal nonlinearly moderates the association between valence and stock market reactions. Specifically, the squared interaction term $\text{AUDIOV} \times \text{AUDIOA}^2$ negatively and significantly relates to the firm's CAR(0,1) after the call ($\beta=-1.063$, $p<0.05$). To further illustrate the moderating effect of AUDIOA, we divide the dataset into three subsamples of equal size with low, medium, and high levels of CEO vocal arousal. The estimation results of Model 1 for the subsamples are provided in Table 4. As shown in the table, the positive association between AUDIOV and CAR(0,1) is statistically significant only in the medium arousal group ($\beta=0.054$, $p<0.05$). Furthermore, Fig. 2 provides a visualization of how CEO vocal valence and arousal interact to affect stock market reactions. Fig. 2(a) plots the regression coefficients (with 95% confidence intervals) of AUDIOV for CAR(0,1) at varying levels of AUDIOA, and Fig. 2 (b) provides a three-dimensional graph to show how CAR(0,1) changes with AUDIOV and AUDIOA when other variables take their fixed mean values. The visualization echoes that the positive association between CEO vocal valence and stock market reactions is strongest at medium levels of arousal, while for too high or too low arousal, the relationship is statistically insignificant at the 0.05 level. Hence, Hypothesis 2 is supported.

In Model 3, we explore the moderating effect of the frequency of emotion reversals by incorporating $\text{AUDIOV} \times \text{REVERSAL}$ in the regression, following Equation (4). As shown by the coefficient of the interaction term, REVERSAL positively and significantly moderates the association between AUDIOV and CAR(0,1) ($\beta=0.233$, $p<0.05$). This finding suggests that a high frequency of emotion reversals can strengthen the positive association between CEO vocal valence and stock market reactions, supporting Hypothesis 3.

Model 4 is a full model that includes all the variables and interaction terms. The coefficients of AUDIOV ($\beta=0.051$), $\text{AUDIOV} \times \text{AUDIOA}^2$ ($\beta=-0.972$), and $\text{AUDIOV} \times \text{REVERSAL}$ ($\beta=0.220$) are all statistically significant, which are consistent with Models 1 to 3. The results further support our proposed hypotheses.

5.3. DiD analysis

To further address the potential endogeneity concerns resulting from simultaneity or unobservable omitted variables (e.g., the CEO's private

Table 2
Descriptive statistics of variables.

Variable	Mean	SD	Min	P25	Median	P75	Max
CAR(0,1)	0.00	0.07	−0.23	−0.04	0.00	0.04	0.23
AUDIOV	0.58	0.10	0.32	0.51	0.58	0.64	0.79
AUDIOA	0.49	0.10	0.24	0.43	0.50	0.56	0.74
REVERSAL	0.24	0.09	0.02	0.18	0.24	0.29	0.47
TEXTV	0.68	0.05	0.54	0.65	0.68	0.72	0.79
TEXTA	0.46	0.02	0.41	0.45	0.46	0.47	0.50
F ₀ (Mean)	5.01	0.15	4.69	4.90	5.00	5.10	5.41
F ₀ (Std)	4.08	0.39	3.24	3.80	4.08	4.37	4.96
JITTER	0.03	0.00	0.02	0.02	0.03	0.03	0.04
SHIMMER	0.14	0.02	0.10	0.13	0.14	0.16	0.19
HNR	2.22	0.23	1.60	2.08	2.24	2.39	2.66
TIME	6.13	0.64	4.11	5.78	6.23	6.58	7.29
PROPORTION	0.47	0.15	0.13	0.37	0.47	0.56	0.88
EXECUTIVES	1.19	0.27	0.69	1.10	1.10	1.39	1.95
ANALYSTS	1.88	0.56	0.00	1.61	1.95	2.30	2.89
ROA	0.02	0.02	−0.05	0.01	0.02	0.03	0.11
MCAP	8.63	1.56	5.59	7.50	8.48	9.68	12.39
GROWTH	0.13	0.35	−0.64	−0.03	0.06	0.20	1.97
BTM	0.42	0.41	−0.24	0.16	0.31	0.56	2.30
LEV	0.24	0.20	0.00	0.09	0.20	0.35	0.82
VOLATILITY	0.03	0.01	0.01	0.02	0.02	0.03	0.08
SURP	0.01	0.05	−0.16	0.00	0.00	0.01	0.33
VOLUME	4.48	8.31	0.08	0.64	1.65	4.22	53.93
GENDER	0.07	0.25	0	0	0	0	1
AGE	4.05	0.11	3.76	3.97	4.06	4.11	4.32
TENURE	1.84	0.80	0.00	1.39	1.79	2.40	3.58

Note: N=10,787; SD: standard deviation; P25: 25th percentile; P75: 75th percentile.

Table 3
Results of two-way fixed effects models.

Variable	Dependent Variable=CAR(0,1)				
	Model 0	Model 1	Model 2	Model 3	Model 4
AUDIOV		0.031*** (0.011)	0.042*** (0.013)	0.042*** (0.012)	0.051*** (0.013)
AUDIOV × AUDIOA ²			−1.063** (0.529)		−0.972* (0.528)
AUDIOV × AUDIOA			0.075 (0.085)		0.045 (0.085)
AUDIOV × REVERSAL				0.233** (0.090)	0.220** (0.092)
AUDIOA ²	−0.051 (0.065)	−0.027 (0.065)	−0.074 (0.070)	−0.035 (0.065)	−0.073 (0.070)
AUDIOA	−0.002 (0.015)	−0.013 (0.016)	−0.009 (0.016)	−0.012 (0.016)	−0.009 (0.016)
REVERSAL	0.007 (0.009)	0.005 (0.009)	0.007 (0.009)	0.004 (0.009)	0.006 (0.009)
Controls	YES	YES	YES	YES	YES
Firm fixed effects	YES	YES	YES	YES	YES
Year–Quarter fixed effects	YES	YES	YES	YES	YES
R ²	0.1398	0.1405	0.1409	0.1412	0.1415
Adj R ²	0.0480	0.0487	0.0489	0.0494	0.0495
Within R ²	0.0353	0.0361	0.0365	0.0368	0.0371
N	10,787	10,787	10,787	10,787	10,787

Note: Robust standard errors are in parentheses; ***, $p < 0.01$; **, $p < 0.05$; *, $p < 0.1$ (two-tailed).

information about firm performance), this study performs a difference-in-differences (DiD) analysis to strengthen the causal interpretation of the relationship between CEO vocal valence and stock market reactions.

5.3.1. Subsample construction

For the DiD analysis, we employ CEO turnover as the exogenous shock. Given the inherent differences in information processing and interpretation across different CEOs, this shock can to some extent enhance the randomness of the independent variable (i.e., CEO vocal valence). To derive the subsample, we first identify firms that meet two

Table 4
Regression results of subsamples with different levels of CEO vocal arousal.

Variable	Dependent Variable=CAR(0,1)		
	Low AUDIOA (Mean=−0.11; SD=0.06)	Medium AUDIOA (Mean=0.00; SD=0.02)	High AUDIOA (Mean=0.11; SD=0.05)
AUDIOV	−0.010 (0.024)	0.054** (0.021)	0.033 (0.020)
AUDIOA ²	−0.287 (0.452)	−1.205 (2.512)	0.324 (0.441)
AUDIOA	−0.087 (0.129)	0.015 (0.065)	−0.085 (0.127)
REVERSAL	−0.087 (0.017)	0.015 (0.016)	−0.085 (0.017)
Controls	YES	YES	YES
Firm fixed effects	YES	YES	YES
Year–Quarter fixed effects	YES	YES	YES
R ²	0.2275	0.3029	0.2556
Adj R ²	0.0391	0.0907	0.0730
Within R ²	0.0472	0.0415	0.0438
N	3,596	3,596	3,595

Note: Robust standard errors are in parentheses; ***, $p < 0.01$; **, $p < 0.05$; *, $p < 0.1$ (two-tailed).

conditions: (1) they experienced a CEO turnover during the sample period, and (2) they maintain complete records within a nine-quarter window centered on the event quarter (four quarters prior and four quarters post). Among them, firms are assigned to the treatment group if the turnover resulted in a significant increase in the average CEO vocal valence, defined as a rise exceeding one-fourth of the standard deviation (i.e., 0.025, as reported in Table 2). Firms experiencing a CEO turnover but with negligible changes in the average AUDIOV (i.e., not exceeding the 0.025 threshold) serve as the control group.

5.3.2. Propensity score matching (PSM)

To mitigate potential selection bias, we employ PSM to construct the treatment and control samples for our DiD analysis (Fang et al., 2025; Meng et al., 2024). The selected firm- and CEO-level matching

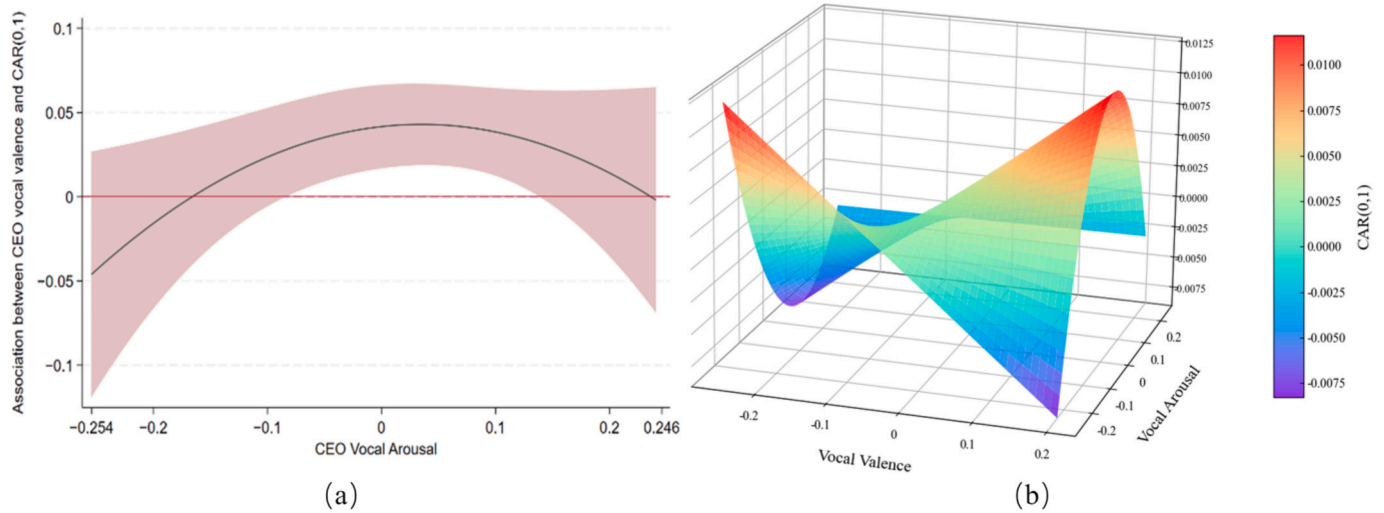


Fig. 2. Visualization of the moderating effect of CEO vocal arousal.

covariates include ROA, MCAP, BTM, AGE, and GENDER, which exhibit significant differences between the two sample groups and are known determinants of stock market reactions (Frankel et al., 2022; Signorello et al., 2020). Following prior studies (Jin & Yu, 2022), we implement a one-to-two nearest neighbor matching protocol with a caliper width of 0.001 to ensure balanced pairwise samples with comparable propensity scores.

Fig. 3 illustrates the propensity score density curves for the treatment and control groups both before and after matching. As depicted in the left panel, significant systematic differences exist between the two groups prior to matching. In contrast, the post-matching density curves in the right panel nearly overlap, indicating that the PSM procedure successfully aligns the treatment and control samples across the observed firm- and CEO-level characteristics. This effectively mitigates the concern of selection bias.

Table 5 presents the balance test results for the matching covariates. Here, “%bias” measures the mean difference of the variable between the two groups, and “%reduct |bias|” represents the reduction in this difference after matching (Rosenbaum & Rubin, 1985). The results show a significant reduction in the absolute bias for all covariates following the PSM. Furthermore, the *t*-tests confirm that no statistically significant differences remain between the matched treatment and control groups across all matching variables. Consequently, the resulting subsample provides a valid basis for the subsequent DiD analysis.

Table 5

Balance tests of covariates before and after matching.

Variable	Unmatched (U)	Matched (M)	%bias	%reduct bias	<i>t</i> -test	
					<i>t</i>	<i>p</i> > <i>t</i>
ROA	U		−11.4%	95.6%	−1.92	0.055
	M		0.5%		0.07	0.944
MCAP	U		−39.5%	90.8%	−5.81	0.000
	M		3.6%		0.86	0.388
BTM	U		11.9%	84.9%	1.97	0.049
	M		−1.8%		−0.24	0.810
AGE	U		13.3%	96.3%	2.13	0.033
	M		−0.5%		−0.07	0.946
GENDER	U		11.9%	19.4%	2.00	0.046
	M		9.6%		1.41	0.159

5.3.3. Parallel trends test

To validate the effectiveness of our exogenous treatment, we conduct a parallel trends test to examine the dynamic association between increased CEO vocal valence in earnings conference calls and stock market reactions. The parallel trends assumption requires that the treatment and control groups exhibit statistically similar trends in CAR (0,1) during the pre-treatment period. Following the common practice in existing research (Rambachan & Roth, 2023), we utilize a series of dummy variables to map the turnover timeline. Specifically, *Current*

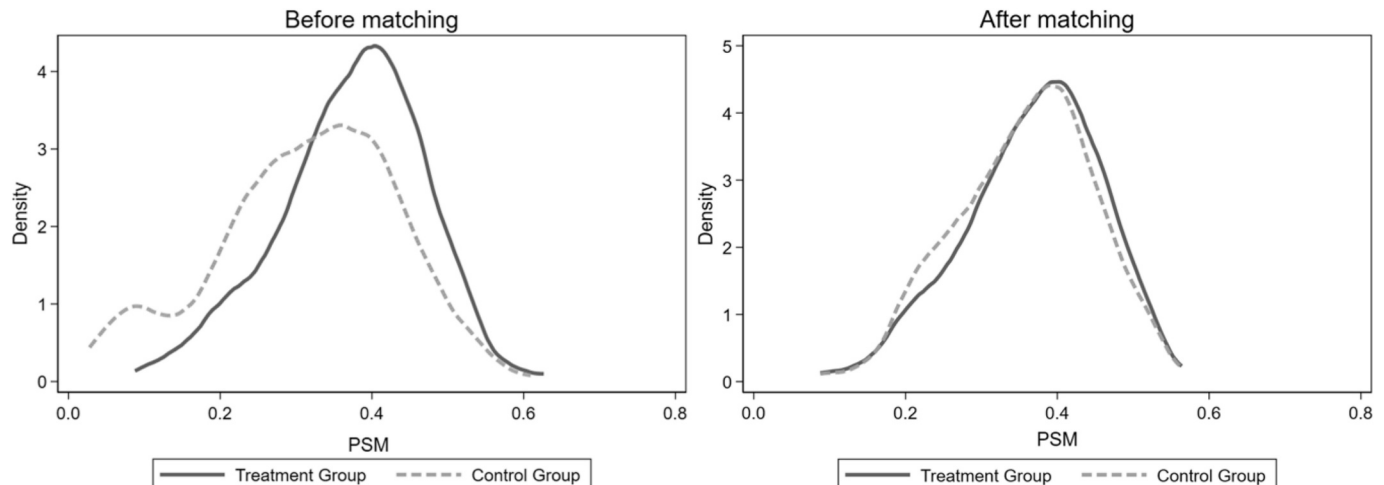


Fig. 3. Density curves of the propensity score before and after matching.

denotes the turnover quarter, *Before*⁻⁴, *Before*⁻³, *Before*⁻², and *Before*⁻¹ represent the four quarters before the turnover, and *After*⁺¹, *After*⁺², *After*⁺³, and *After*⁺⁴ represent the quarters in the post-turnover period. Using *Before*⁻⁴ as the benchmark, Table 6 presents the two-way fixed effects regression results of the parallel trends test.

As shown in Table 6, the pre-turnover coefficients are all statistically insignificant at the 0.1 level, confirming that no systematic differences in CAR(0,1) exist between the two groups prior to the shock. However, the estimated coefficients become positive and statistically significant (except for *After*⁺¹) following the CEO turnover, revealing a persistent gap in stock market reactions. Specifically, the treatment group exhibits significantly higher CAR(0,1) compared to the control group. These results generally satisfy the parallel trends assumption and validate the use of CEO turnover as an exogenous shock, providing further evidence on CEO vocal valence's relationship with stock market reactions.

5.3.4. DiD model estimation

Using the matched treatment and control samples, we estimate the DiD model specified in Equation (5).

$$CAR(0,1)_{i,t} = \beta_0 + \beta_8 Treat_i \times Post_{i,t} + \beta_2 AUDIOA_{i,t} + \beta_3 REVERSAL_{i,t} + \beta_4 AUDIOA_{i,t}^2 + \alpha X_{i,t} + \gamma_i + \delta_t + \varepsilon_{i,t}, \quad (5)$$

where *Treat*_{*i*} is a dummy variable equal to one if firm *i* belongs to the treatment group and zero otherwise, and *Post*_{*i,t*} is a dummy variable equal to one for firm-quarter observations occurring after the CEO turnover. Consistent with Hypothesis 1, the coefficient β_8 is expected to be positive, suggesting that a rise in CEO vocal valence is associated with an increase in stock market reactions.

The PSM-DiD estimation results are summarized in Table 7. Column (0) reports the baseline estimation for Equation (5). The coefficient for the interaction term *Treat*×*Post* is positive ($\beta=0.025$) and statistically significant at the 0.05 level, indicating that firms experience a significant increase in CAR(0,1) following a transition to a CEO with higher average vocal valence in earnings conference calls. To ensure the robustness of this finding, we perform several sensitivity analyses. Specifically, we adjust the vocal valence increase threshold to one-third and one-half of the standard deviation (i.e., columns (1) and (2), respectively) and shorten the observation window to seven and five quarters (i.e., columns (3) and (4), respectively) before re-conducting the PSM. As shown in Table 7, the DiD estimates remain positive and significant across all specifications, further supporting Hypothesis 1; that is, CEO vocal valence in earnings conference calls is positively associated with stock market reactions.

Table 6
Regression results of the parallel trends test.

Variable	Dependent Variable=CAR(0,1)	
<i>Before</i> ⁻³	0.017	(0.018)
<i>Before</i> ⁻²	0.015	(0.018)
<i>Before</i> ⁻¹	0.029	(0.019)
Current	0.046**	(0.019)
<i>After</i> ⁺¹	0.031	(0.023)
<i>After</i> ⁺²	0.047**	(0.023)
<i>After</i> ⁺³	0.041*	(0.024)
<i>After</i> ⁺⁴	0.061**	(0.026)
Controls	YES	
Firm fixed effects	YES	
Year–Quarter fixed effects	YES	
R ²	0.3153	
Adj R ²	0.0434	
Within R ²	0.0501	
N	754	

Notes: Robust standard errors are in parentheses; ***, $p < 0.01$; **, $p < 0.05$; *, $p < 0.1$ (two-tailed).

Table 7
Results of DiD models.

Variable	Dependent Variable=CAR(0,1)				
	(0)	(1)	(2)	(3)	(4)
<i>Treat</i> × <i>Post</i>	0.025** (0.012)	0.027* (0.014)	0.036** (0.014)	0.023* (0.014)	0.032** (0.014)
Controls	YES	YES	YES	YES	YES
Firm fixed effects	YES	YES	YES	YES	YES
Year–Quarter fixed effects	YES	YES	YES	YES	YES
R ²	0.3104	0.2866	0.3521	0.3918	0.5256
Adj R ²	0.0490	0.0082	0.0643	0.0840	0.1612
Within R ²	0.0434	0.0481	0.0693	0.0505	0.1141
N	754	742	687	635	482

Notes: Robust standard errors are in parentheses; ***, $p < 0.01$; **, $p < 0.05$; *, $p < 0.1$ (two-tailed).

5.3.5. Additional analyses

To ensure the robustness of our baseline findings, we address potential concerns arising from heterogeneous treatment effects in staggered DiD settings. When CEO turnover timing varies across firms, standard two-way fixed effects estimators may leverage previously treated firms that experienced a significant increase in the average CEO vocal valence due to earlier turnover as controls. If treatment effects linger or evolve over time, such comparisons may generate negative weights and distort coefficient estimates (Baker et al., 2022; de Chaisemartin & d'Haultfoeuille, 2020; Sun & Abraham, 2021). Therefore, we re-estimate our model using two alternative state-of-the-art estimators that rely exclusively on “clean” controls (i.e., never-treated firms). That is, the stacked DiD estimator uniquely suited to short-window corporate event studies (Bleiberg, 2021; Cengiz et al., 2019), and the interaction-weighted estimator for tracing dynamic treatment effects (Sun & Abraham, 2021). Specifically, the stacked DiD setup conducts cohort-specific analyses, where each cohort consists of firms experiencing CEO turnover with a significant increase in vocal valence in a specific year–quarter, matched with contemporaneous never-treated control firms. The interaction-weighted estimation provides interpretable period-by-period coefficients, where *T*=0 denotes the quarter of CEO

Table 8
DiD results with alternative estimators.

Variable	Dependent Variable=CAR(0,1)	
	Stacked DiD estimator	Interaction-weighted estimator
<i>Treat</i> × <i>Post</i>	0.043*** (0.016)	
Dynamic post-treatment estimates		
<i>T</i> =0		0.045** (0.019)
<i>T</i> =1		0.031 (0.023)
<i>T</i> =2		0.039* (0.021)
<i>T</i> =3		0.049** (0.021)
<i>T</i> =4		0.059** (0.025)
Controls	YES	YES
Cohort×Firm fixed effects	YES	
Cohort×Year–Quarter fixed effects	YES	
Firm fixed effects		YES
Year–Quarter fixed effects		YES
R ²	0.3806	0.4130
Adj R ²	0.1263	0.3843
Within R ²	0.0722	0.1047
N	6,432	6,522

Note: Robust standard errors are in parentheses; ***, $p < 0.01$; **, $p < 0.05$; *, $p < 0.1$ (two-tailed).

turnover, while T=1, 2, 3, and 4 represent the four subsequent post-event quarters.

The re-estimation results are reported in Table 8. Both estimators yield positive and statistically significant treatment effects that are highly consistent with our baseline estimates. The stacked DiD coefficient for Treat×Post is 0.043, which is significant at the 0.01 level. The interaction-weighted estimator further confirms a persistently positive stock market reaction (i.e., increase in CAR(0,1)) across post-event periods, with coefficients remaining statistically significant in most horizons. Overall, these results demonstrate that our core conclusion is robust and not materially distorted by negative weighting or treatment effect heterogeneity.

Furthermore, to test whether the moderating effects of CEO vocal arousal (Hypothesis 2) and the frequency of emotion reversals (Hypothesis 3) hold under a quasi-experimental setup, we extend our baseline DiD model in Equation (5) by introducing interacted treatment terms (i.e., $Treat_{it} \times Post_{it} \times Moderator_{it}$). The empirical results are presented in Table 9.

As shown in column (3) of Table 9, the triple-interaction term coefficient for the frequency of emotion reversals (Treat×Post×REVERSAL) is positive ($\beta=0.178$) and statistically significant at the 0.01 level, confirming that the moderating role of REVERSAL remains robust. However, as shown in column (2), the interaction term for the quadratic CEO vocal arousal (Treat×Post×AUDIOA²) does not achieve statistical significance ($\beta=-0.296$, $p>0.1$). We attribute this divergence primarily to statistical power limitations and the identification design. First, strict PSM procedures substantially reduce our sample from 10,787 panel observations to 754 matched observations. Detecting non-linear and high-order interactions on such a restricted sample inherently suffers from reduced statistical power and inflated standard errors. Second, our identification strategy utilizes CEO turnover as an exogenous shock specifically tailored to isolate shifts in vocal valence. Because the matching procedure is optimized around this valence shock, the matched control group may not serve as an ideal counterfactual across heterogeneous levels of vocal arousal. Consequently, while the PSM-DiD setting successfully validates the main association proposed in Hypothesis 1 and the robustness of Hypothesis 3, this study retains full-panel two-way fixed effects models as the primary specification to ensure adequate statistical power for capturing complex and non-linear moderating effects in the complete empirical framework.

Table 9
DiD results with moderating effects.

Variable	Dependent Variable=CAR(0,1)		
	(1)	(2)	(3)
Treat×Post	0.025** (0.012)	-0.027 (0.227)	-0.022 (0.023)
Treat×Post × AUDIOA ²		-0.296 (0.745)	
Treat×Post × AUDIOA		0.258 (0.829)	
Treat×Post × REVERSAL			0.178*** (0.068)
AUDIOA ²	0.094 (0.282)	0.142 (0.291)	0.066 (0.285)
AUDIOA	-0.155 (0.284)	-0.195 (0.294)	-0.126 (0.286)
REVERSAL	0.033 (0.033)	0.036 (0.033)	-0.012 (0.035)
Controls	YES	YES	YES
Firm fixed effects	YES	YES	YES
Year-Quarter fixed effects	YES	YES	YES
R ²	0.3104	0.3108	0.3155
Adj R ²	0.0490	0.0495	0.0578
Within R ²	0.0434	0.0438	0.0505
N	754	754	754

Note: Robust standard errors are in parentheses; ***, $p < 0.01$; **, $p < 0.05$; *, $p < 0.1$ (two-tailed).

5.4. Mechanism analysis

To substantiate the investor cognitive processing story, this subsection conducts mechanism tests addressing two core questions: (1) Does CEO vocal emotion shape stock market reactions by altering investors' cognitive attention? and (2) How investors dynamically process and react to CEO vocal signals over time?

Under the cognitive-processing framework, positive nonverbal vocal cues activate the appetitive motivational system, raising market participants' attentiveness during earnings conference calls (Wang & Lang, 2012). To test this prediction, we follow prior literature and proxy for investor attention using stock turnover over the call window (TURN-OVER(0,1)), computed as trading volume divided by total shares outstanding (Yu et al., 2026). Extant research establishes that when investors allocate greater cognitive resources to an information event, they trade more actively, producing higher stock turnover (Gargano & Rossi, 2018; Loh, 2010). In addition, we capture analyst engagement and call interactivity using the natural logarithm of the number of analyst questions during the Q&A session (QUESTION), which reflects the cognitive processing intensity of sophisticated market participants (Matsumoto et al., 2011).

As reported in column (1) of Table 10, CEO vocal valence is positively and significantly associated with stock turnover ($\beta=0.179$, $p<0.01$). This confirms that positive vocal tone actively captures investors' cognitive attention and elevates trading intensity. Column (2) shows that AUDIOV also exhibits a positive and statistically significant relationship with the number of analyst questions ($\beta=0.654$, $p<0.01$), indicating that CEOs' positive nonverbal cues in earnings conference calls encourage active inquiry and participation among market professionals.

To further examine how market responses evolve as CEO vocal signals are processed, we decompose cumulative abnormal returns and raw stock returns into call-day contemporaneous reactions (AR(0), IR(0)) and post-call day reactions (AR(1), IR(1)). The estimation results in Table 11 reveal a temporal divergence driven by signal complexity.

On one hand, investors display immediate reactions to primary AUDIOV cues. As shown in the table, the direct effect of CEO vocal valence is concentrated almost entirely on the call day and becomes statistically insignificant on the subsequent trading day. Because vocal valence functions as a primary and intuitively decoded sentiment signal, investors process it with minimal cognitive friction, resulting in rapid price discovery on Day 0. On the other hand, investors exhibit lagged responses to more complex signals such as AUDIOA and REVERSAL. In contrast to the immediate direct effect, the moderating effects of CEO vocal arousal and the frequency of emotion reversals are statistically insignificant on Day 0, but become pronounced on Day 1. This temporal separation stems from differences in signal disambiguation costs. Unlike valence, arousal and reversals are multifaceted signals. Intense vocal arousal can signify either genuine managerial confidence or stress-

Table 10
Mechanism analysis: investor attention.

Variable	Dependent Variable=TURNOVER(0,1)	Dependent Variable=QUESTION
	(1)	(2)
AUDIOV	0.179*** (0.064)	0.654*** (0.044)
Controls	YES	YES
Firm fixed effects	YES	YES
Year-Quarter fixed effects	YES	YES
R ²	0.7810	0.7191
Adj R ²	0.7576	0.6891
Within R ²	0.2889	0.3457
N	10,787	10,787

Note: Robust standard errors are in parentheses; ***, $p < 0.01$; **, $p < 0.05$; *, $p < 0.1$ (two-tailed).

Table 11
Mechanism analysis: temporal dynamics of returns.

Variable	Dependent Variable=AR(0)			Dependent Variable=AR(1)		
	(1)	(2)	(3)	(4)	(5)	(6)
AUDIOV	0.031*** (0.011)	0.033** (0.013)	0.037*** (0.012)	0.004 (0.005)	0.010* (0.006)	0.009* (0.005)
AUDIOV × AUDIOA ²		−0.343 (0.538)			−0.562** (0.254)	
AUDIOV × AUDIOA		0.124 (0.083)			−0.028 (0.038)	
AUDIOV × REVERSAL			0.128 (0.090)			0.116*** (0.039)
AUDIOA ²	−0.020 (0.066)	−0.054 (0.071)	−0.025 (0.066)	0.008 (0.032)	−0.004 (0.032)	0.003 (0.032)
AUDIOA	−0.028* (0.015)	−0.025 (0.015)	−0.028* (0.015)	0.008 (0.007)	0.009 (0.007)	0.008 (0.007)
REVERSAL	−0.004 (0.009)	−0.002 (0.009)	−0.004 (0.009)	0.009** (0.004)	0.009** (0.004)	0.008** (0.004)
Controls	YES	YES	YES	YES	YES	YES
Firm fixed effects	YES	YES	YES	YES	YES	YES
Year–Quarter fixed effects	YES	YES	YES	YES	YES	YES
R ²	0.1400	0.1402	0.1402	0.1159	0.1164	0.1167
Adj R ²	0.0482	0.0482	0.0483	0.0214	0.0218	0.0222
Within R ²	0.0351	0.0353	0.0353	0.0078	0.0084	0.0087
N	10,787	10,787	10,787	10,787	10,787	10,787
Variable	Dependent Variable=IR(0)			Dependent Variable=IR(1)		
	(7)	(8)	(9)	(10)	(11)	(12)
AUDIOV	0.033*** (0.011)	0.035*** (0.013)	0.038*** (0.012)	0.006 (0.005)	0.012* (0.006)	0.010* (0.006)
AUDIOV × AUDIOA ²		−0.337 (0.555)			−0.540* (0.292)	
AUDIOV × AUDIOA		0.095 (0.086)			−0.058 (0.043)	
AUDIOV × REVERSAL			0.107 (0.091)			0.094** (0.044)
AUDIOA ²	−0.057 (0.067)	−0.086 (0.072)	−0.061 (0.067)	−0.003 (0.037)	−0.008 (0.037)	−0.007 (0.037)
AUDIOA	−0.031** (0.015)	−0.028* (0.015)	−0.030** (0.015)	0.016** (0.008)	0.016** (0.008)	0.016** (0.008)
REVERSAL	−0.004 (0.009)	−0.002 (0.009)	−0.004 (0.009)	0.009* (0.005)	0.009* (0.005)	0.008* (0.005)
Controls	YES	YES	YES	YES	YES	YES
Firm fixed effects	YES	YES	YES	YES	YES	YES
Year–Quarter fixed effects	YES	YES	YES	YES	YES	YES
R ²	0.1424	0.1425	0.1425	0.1196	0.1201	0.1201
Adj R ²	0.0507	0.0507	0.0508	0.0256	0.0260	0.0260
Within R ²	0.0341	0.0342	0.0342	0.0071	0.0077	0.0076
N	10,787	10,787	10,787	10,787	10,787	10,787

Note: Robust standard errors are in parentheses; ***, $p < 0.01$; **, $p < 0.05$; *, $p < 0.1$ (two-tailed).

induced anxiety (Yin et al., 2017); similarly, emotion reversals convey fine-grained dynamic fluctuations throughout the call that require context-dependent interpretation (Nabi & Green, 2015). Given the cognitive capacity constraints during live earnings conference calls, investors cannot fully disambiguate these complex nonverbal signals in real time (DellaVigna & Pollet, 2009). Full absorption requires post-call cognitive processing, such as reviewing transcripts, consulting commentary, and re-evaluating context (Blankespoor et al., 2020; Lyle et al., 2025). Consequently, the market price reaction for complex vocal cues materializes on Day 1 as investors complete their information synthesis.

Taken together, these empirical findings provide multi-angled support for the investor cognitive processing mechanism. By demonstrating both the activation of cognitive attention and the differentiated processing speeds associated with varying signal complexities, the results confirm that CEO vocal emotion shapes stock market reactions by modulating investor cognitive allocation and information processing.

5.5. Robustness checks

5.5.1. Alternative recording sections

In the above analysis, CEO vocal emotion is measured with the

MD&A section of earnings conference call recordings. For robustness checks, we extend the scope to the entire call including the question and answer (Q&A) section, which is naturally more interactive and spontaneous (Rennekamp et al., 2022). Based on the recalculated variables (i. e., AUDIOV, AUDIOA, REVERSAL, TEXTV, TEXTA, F0(Mean), F0(Std), JITTER, SHIMMER, HNR, TIME, PROPORTION) corresponding to the entire recordings, Table 12 presents the re-estimated results of the regression models. As shown in the table, the coefficients of focus and their statistical significance validate the robustness of our hypothesis testing conclusions for entire calls.

5.5.2. Alternative measures of the dependent and independent variables

As shown in Table 1, the dependent variable CAR(0,1) is calculated using the S&P 1500 index. To check robustness, we adopt another common calculation of CAR(0,1) based on corporate daily returns from the CRSP database (Dong & Wu, 2015; Yu & Leistikow, 2011). Columns (1) to (3) of Table 13 show the regression results. In addition, we perform gender-grouped z-score normalization and then scale the normalized values to the range of [0,1] to construct an alternative measure of the independent variable (AUDIOV), and the re-estimation results are provided in columns (4) to (6) of Table 13. The results

Table 12

Robustness checks: alternative recording sections.

Variable	Dependent Variable=CAR(0,1)		
	(1)	(2)	(3)
AUDIOV	0.035*** (0.010)	0.043*** (0.011)	0.041*** (0.011)
AUDIOV × AUDIOA ²		−0.758* (0.459)	
AUDIOV × AUDIOA		0.049 (0.079)	
AUDIOV × REVERSAL			0.253** (0.121)
AUDIOA ²	0.020 (0.065)	−0.018 (0.071)	0.012 (0.065)
AUDIOA	−0.017 (0.016)	−0.014 (0.016)	−0.016 (0.016)
REVERSAL	0.012 (0.015)	0.014 (0.015)	0.012 (0.015)
Controls	YES	YES	YES
Firm fixed effects	YES	YES	YES
Year–Quarter fixed effects	YES	YES	YES
R ²	0.1442	0.1444	0.1446
Adj R ²	0.0492	0.0493	0.0497
Within R ²	0.0374	0.0376	0.0379
N	10,429	10,429	10,429

Note: Robust standard errors are in parentheses; ***, $p < 0.01$; **, $p < 0.05$; *, $p < 0.1$ (two-tailed).

consistently support the proposed three hypotheses. Specifically, CEO vocal valence in earnings conference calls is positively associated with stock market reactions, and CEO vocal arousal and the frequency of emotion reversals significantly moderate this association.

5.5.3. Alternative measures of the moderator variables

To test the robustness of the moderating effects, we adopt alternative measures of CEO vocal arousal and the frequency of emotion reversals to re-estimate the regression models in Equations (3) and (4). Specifically, for AUDIOA, we scale vocal arousal scores to the range of [0,1] by gender group, and the re-estimation results are shown in column (1) of Table 14. Corresponding to column (2), another alternative measure of AUDIOA is constructed by performing gender-grouped z-score normalization before range scaling. For emotion reversals, the threshold of valence shift is changed to one and three standard deviations (i.e., 0.1

and 0.3), and the regression results are provided in columns (3) and (4), respectively. As indicated in Table 14, the conclusions for Hypotheses 2 and 3 remain robust across different measures of the moderator variables, implicating that CEO vocal arousal nonlinearly moderates the association between AUDIOV and CAR(0,1), and the positive association is strongest at medium levels of arousal. Moreover, a high frequency of emotion reversals strengthens the positive association between CEO vocal valence and stock market reactions.

5.5.4. Additional control variables

As elaborated in Section 4.2.4, to isolate the association between CEO vocal emotion and stock market reactions, this study controls for several acoustic features in the model specification. In previous research on CEO vocal emotion, machine learning methods have also been commonly used to identify discrete emotional states in audio data and

Table 14

Robustness checks: alternative measures of moderator variables.

Variable	Dependent Variable=CAR(0,1)			
	(1)	(2)	(3)	(4)
AUDIOV	0.042*** (0.013)	0.042*** (0.013)	0.043*** (0.013)	0.035*** (0.011)
AUDIOV × AUDIOA ²	−0.265** (0.132)	−0.345** (0.167)		
AUDIOV × AUDIOA	0.037 (0.042)	0.041 (0.048)		
AUDIOV × REVERSAL			0.147** (0.075)	0.191* (0.111)
AUDIOA ²	−0.018 (0.017)	−0.021 (0.022)	−0.033 (0.066)	−0.033 (0.065)
AUDIOA	−0.016 (0.028)	−0.016 (0.028)	−0.013 (0.016)	−0.012 (0.016)
REVERSAL	0.007 (0.009)	0.007 (0.009)	0.002 (0.008)	0.004 (0.012)
Controls	YES	YES	YES	YES
Firm fixed effects	YES	YES	YES	YES
Year–Quarter fixed effects	YES	YES	YES	YES
R ²	0.1409	0.1409	0.1409	0.1408
Adj R ²	0.0489	0.0489	0.0490	0.0489
Within R ²	0.0365	0.0365	0.0365	0.0364
N	10,787	10,787	10,787	10,787

Note: Robust standard errors are in parentheses; ***, $p < 0.01$; **, $p < 0.05$; *, $p < 0.1$ (two-tailed).

Table 13

Robustness checks: alternative measures of dependent and independent variables.

Variable	Dependent Variable=CAR(0,1)					
	(1)	(2)	(3)	(4)	(5)	(6)
AUDIOV	0.031*** (0.011)	0.042*** (0.013)	0.042*** (0.012)	0.017*** (0.006)	0.023*** (0.007)	0.022*** (0.006)
AUDIOV × AUDIOA ²		−1.071** (0.527)			−0.585** (0.284)	
AUDIOV × AUDIOA		0.077 (0.085)			0.032 (0.046)	
AUDIOV × REVERSAL			0.237*** (0.091)			0.132*** (0.048)
AUDIOA ²	−0.026 (0.065)	−0.074 (0.070)	−0.035 (0.065)	−0.027 (0.065)	0.209 (0.129)	−0.036 (0.065)
AUDIOA	−0.014 (0.016)	−0.010 (0.016)	−0.013 (0.016)	−0.013 (0.016)	−0.025 (0.026)	−0.012 (0.016)
REVERSAL	0.006 (0.009)	0.008 (0.009)	0.005 (0.009)	0.005 (0.009)	0.007 (0.009)	−0.059** (0.025)
Controls	YES	YES	YES	YES	YES	YES
Firm fixed effects	YES	YES	YES	YES	YES	YES
Year–Quarter fixed effects	YES	YES	YES	YES	YES	YES
R ²	0.1405	0.1408	0.1412	0.1405	0.1409	0.1413
Adj R ²	0.0487	0.0489	0.0493	0.0487	0.0489	0.0495
Within R ²	0.0358	0.0362	0.0366	0.0361	0.0365	0.0369
N	10,787	10,787	10,787	10,787	10,787	10,787

Note: Robust standard errors are in parentheses; ***, $p < 0.01$; **, $p < 0.05$; *, $p < 0.1$ (two-tailed).

further explore their influence or predictive efficacy (Ewertz et al., 2025; Yang et al., 2023). Therefore, we employ the high-precision SenseVoiceSmall model (An et al., 2024) to extract seven discrete emotional states (i.e., happy, sad, angry, neutral, fearful, disgusted, and surprised) from CEO speech in the MD&A section, which are classified as positive, neutral, and negative (Chen et al., 2023; Momtaz, 2021), and incorporate them as additional control variables in the two-way fixed effects models for robustness testing. The results are presented in Table 15. From the re-estimated coefficients and statistical significance, it can be observed that all three hypotheses proposed in this study continue to be supported.

5.5.5. Alternative fixed effects

For robustness check purposes, we also use alternative fixed effects in regression models. Considering that CEOs' vocal features change with age, which can affect auditory perceptual evaluations and psychosocial attributions (Lortie et al., 2018; Rojas et al., 2020), we incorporate age–quarter fixed effects to control for age-specific influences that may not be fully captured by linear specification in the regression. The corresponding results are presented in columns (1) to (3) in Table 16. Furthermore, since 29% of the CEOs in our sample have experienced a turnover event, we include CEO fixed effects in regression models to control for unobserved CEO-specific characteristics (e.g., charisma, leadership style, general communication ability), following Conyon et al. (2019). The re-estimation results are shown in columns (4) to (6) in Table 16. The estimated coefficients and their statistical significance demonstrate that our hypothesis testing conclusions are robust when the alternative fixed effects are considered.

6. Discussion and conclusion

CEO vocal emotion in earnings conference calls acts as important nonverbal cues to inform investors about firm prospects and relate to stock market reactions. Previous studies encounter limitations in characterizing the interplay between different vocal emotional dimensions, incorporating fine-grained fluctuations in the dynamic emotional flow, and exploring the potential mechanism and boundary conditions. To address existing issues, this study investigates the relationship between CEO vocal emotion and stock market reactions from the perspective of investor cognitive processing. Specifically, using a pretrained SER model

to measure CEO vocal emotion automatically and rapidly, we explore the association between CEO vocal valence in earnings conference calls and stock market reactions and the moderating effects of CEO vocal arousal and the frequency of emotion reversals. Empirical tests and multi-view robustness checks are conducted with two-way fixed effects models and DiD analysis on a large dataset. The results show that CEO vocal valence is positively associated with stock market reactions (i.e., measured by CAR(0,1)). Furthermore, CEO vocal arousal nonlinearly moderates this association, which is strongest at medium levels of arousal; a high frequency of emotion reversals strengthens the positive association between CEO vocal valence and stock market reactions.

This work can deepen the understanding of the association between CEO vocal emotion in earnings conference calls and stock market reactions and provide managerial implications for corporate information disclosure practices. From a theoretical viewpoint, first, unlike prior work that predominately focuses on CEO vocal valence, this study employs SER techniques to measure CEO vocal emotion during earnings conference calls in a 2D continuous space, so as to investigate how valence and arousal nonlinearly interact to influence stock market reactions. Second, in contrast to research efforts that only calculate the aggregated or average vocal emotion, this study considers CEO emotion reversals in the dynamic emotional flow, which provides fine-grained insights into the role of CEO vocal valence. Third, this study broadens the scope of previous studies by revealing the underlying mechanism, that is, CEO vocal valence in earnings conference calls activates investors' motivational systems and shapes cognitive processing, thereby influencing their investment decisions. The validated moderating effects identify the boundary conditions of the association between CEO vocal emotion and stock market reactions and provide evidence for our mechanism exploration from the perspective of investor cognitive processing.

From a practical viewpoint, we conduct an in-depth analysis of the role of CEO vocal emotion in earnings conference calls, which highlights the importance of nonverbal cues in market communication. While financial performance metrics and verbal content have long been the focus in financial market practices, executive vocal emotion acts as a critical yet underexplored dimension of corporate voluntary disclosures. The findings of our study can provide corporations with valuable insights to disclose firm information with suitable tactics such as medium vocal arousal and frequent emotion reversals, which is beneficial for increasing the cognitive processing and engagement of investors, and thus amplifying positive market reactions generated from high vocal valence of executives. Moreover, the conclusions can raise awareness among regulators, analysts, and the media about the potential role of executive emotional cues in earnings conference calls in shaping market responses.

This study has several limitations. First, while our empirical design incorporates extensive control variables, two-way fixed effects, and a PSM-DiD identification test, the observational nature of market data means the findings are best interpreted as robust associations rather than definitive causal effects. Future research could utilize experimental setups or unique exogenous shocks to further isolate these causal mechanisms. Second, our study relies on a pretrained SER model to capture CEO vocal emotion. Future research may employ more advanced methods (e.g., LLM) to improve the measurement or extend the quantification to the three-dimensional space. Third, our primary focus on the emotion reversal frequency may overlook micro-level nuances, so future work could explore more granular features within emotional flows to deepen the understanding of the dynamic association between CEO vocal emotion and stock market reactions. Fourth, our empirical framework isolates the role of CEO vocal cues. Based on the multimodal data, future work may further broaden the research scope to investigate the complex interplay of verbal and nonverbal cues during earnings conference calls and its potential relationship with investors' decision-making.

Table 15
Robustness checks: additional control variables.

Variable	Dependent Variable=CAR(0,1)		
	(1)	(2)	(3)
AUDIOV	0.031*** (0.011)	0.042*** (0.013)	0.042*** (0.012)
AUDIOV × AUDIOA ²		−1.063** (0.528)	
AUDIOV × AUDIOA		0.073 (0.085)	
AUDIOV × REVERSAL			0.236*** (0.091)
AUDIOA ²	−0.027 (0.067)	−0.074 (0.071)	−0.036 (0.067)
AUDIOA	−0.014 (0.016)	−0.010 (0.016)	−0.013 (0.016)
REVERSAL	0.005 (0.009)	0.007 (0.010)	0.004 (0.009)
Controls	YES	YES	YES
Firm fixed effects	YES	YES	YES
Year–Quarter fixed effects	YES	YES	YES
R ²	0.1407	0.1411	0.1414
Adj R ²	0.0486	0.0488	0.0493
Within R ²	0.0361	0.0365	0.0369
N	10,786	10,786	10,786

Note: Robust standard errors are in parentheses; ***, $p < 0.01$; **, $p < 0.05$; *, $p < 0.1$ (two-tailed).

Table 16

Robustness checks: alternative fixed effects.

Variable	Dependent Variable=CAR(0,1)					
	(1)	(2)	(3)	(4)	(5)	(6)
AUDIOV	0.033*** (0.011)	0.043*** (0.013)	0.042*** (0.012)	0.036*** (0.012)	0.047*** (0.014)	0.046*** (0.013)
AUDIOV × AUDIOA ²		−1.011* (0.522)			−1.055* (0.598)	
AUDIOV × AUDIOA		0.054 (0.084)			0.061 (0.095)	
AUDIOV × REVERSAL			0.214** (0.090)			0.212** (0.097)
AUDIOA ²	−0.016 (0.066)	−0.057 (0.070)	−0.024 (0.066)	−0.011 (0.079)	−0.057 (0.084)	−0.020 (0.079)
AUDIOA	−0.004 (0.016)	−0.001 (0.016)	−0.003 (0.016)	0.005 (0.019)	0.009 (0.019)	0.006 (0.019)
REVERSAL	0.004 (0.009)	0.006 (0.009)	0.003 (0.009)	0.006 (0.010)	0.008 (0.010)	0.005 (0.010)
Controls	YES	YES	YES	YES	YES	YES
Firm fixed effects	YES	YES	YES			
Age–Quarter fixed effects	YES	YES	YES			
CEO fixed effects				YES	YES	YES
Year–Quarter fixed effects				YES	YES	YES
R ²	0.1379	0.1382	0.1385	0.1782	0.1785	0.1787
Adj R ²	0.0443	0.0444	0.0448	0.0576	0.0578	0.0581
Within R ²	0.0338	0.0341	0.0344	0.0380	0.0384	0.0386
N	10,787	10,787	10,787	10,787	10,787	10,787

Note: Robust standard errors are in parentheses; ***, $p < 0.01$; **, $p < 0.05$; *, $p < 0.1$ (two-tailed).**CRedit authorship contribution statement**

Baojun Ma: Supervision, Project administration, Methodology, Funding acquisition, Conceptualization. **Lingyun Zhou:** Writing – original draft, Visualization, Validation, Methodology, Investigation. **Yao Mu:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Yi Chen:** Writing – review & editing, Supervision, Methodology, Formal analysis. **Jian Zhang:** Supervision, Methodology, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by National Natural Science Foundation of China [grant numbers 72172092, 72302153, 72672117] and Shanghai Chenguang Program [grant number 24cga34].

Appendix**Table A1**

Full regression results of two-way fixed effects models.

Variable	Dependent Variable=CAR(0,1)				
	Model 0	Model 1	Model 2	Model 3	Model 4
AUDIOV		0.031*** (0.011)	0.042*** (0.013)	0.042*** (0.012)	0.051*** (0.013)
AUDIOV × AUDIOA ²			−1.063** (0.529)		−0.972* (0.528)
AUDIOV × AUDIOA			0.075 (0.085)		0.045 (0.085)
AUDIOV × REVERSAL				0.233** (0.090)	0.220** (0.092)
AUDIOA ²	−0.051 (0.065)	−0.027 (0.065)	−0.074 (0.065)	−0.035 (0.065)	−0.073 (0.070)
AUDIOA	−0.002 (0.015)	−0.013 (0.016)	−0.009 (0.016)	−0.012 (0.016)	−0.009 (0.016)
REVERSAL	0.007 (0.009)	0.005 (0.009)	0.007 (0.009)	0.004 (0.009)	0.006 (0.009)
TEXTV	0.165*** (0.022)	0.149*** (0.023)	0.150*** (0.023)	0.149*** (0.023)	0.150*** (0.023)
TEXTA	0.005 (0.063)	0.005 (0.063)	0.004 (0.063)	0.004 (0.063)	0.003 (0.063)
F ₀ (Mean)	0.003	0.010	0.010	0.010	0.010

(continued on next page)

Table A1 (continued)

Variable	Dependent Variable=CAR(0,1)				
	Model 0	Model 1	Model 2	Model 3	Model 4
F ₀ (Std)	(0.012) 0.001 (0.004)	(0.012) −0.000 (0.004)	(0.012) −0.000 (0.004)	(0.012) −0.000 (0.004)	(0.012) −0.000 (0.004)
JITTER	0.135 (0.319)	0.121 (0.321)	0.111 (0.321)	0.110 (0.320)	0.099 (0.320)
SHIMMER	0.001 (0.063)	0.014 (0.063)	0.014 (0.063)	0.013 (0.063)	0.013 (0.063)
HNR	−0.005 (0.006)	−0.006 (0.006)	−0.006 (0.006)	−0.006 (0.006)	−0.006 (0.006)
TIME	−0.002 (0.002)	−0.001 (0.002)	−0.001 (0.002)	−0.001 (0.002)	−0.001 (0.002)
PROPORTION	0.005 (0.010)	0.005 (0.010)	0.005 (0.010)	0.005 (0.010)	0.005 (0.010)
EXECUTIVES	−0.002 (0.006)	−0.002 (0.006)	−0.002 (0.006)	−0.002 (0.006)	−0.002 (0.006)
ANALYSTS	0.008*** (0.003)	0.009*** (0.003)	0.009*** (0.003)	0.009*** (0.003)	0.008*** (0.003)
ROA	0.439*** (0.063)	0.440*** (0.063)	0.440*** (0.063)	0.441*** (0.063)	0.441*** (0.063)
MCAP	−0.042*** (0.004)	−0.042*** (0.004)	−0.042*** (0.004)	−0.042*** (0.004)	−0.042*** (0.004)
GROWTH	−0.001 (0.003)	−0.001 (0.003)	−0.001 (0.003)	−0.001 (0.003)	−0.001 (0.003)
BTM	0.007 (0.007)	0.007 (0.007)	0.007 (0.007)	0.007 (0.007)	0.007 (0.007)
LEV	−0.024 (0.019)	−0.024 (0.019)	−0.023 (0.019)	−0.023 (0.019)	−0.023 (0.019)
VOLATILITY	−0.099 (0.109)	−0.089 (0.109)	−0.089 (0.109)	−0.087 (0.108)	−0.087 (0.108)
SURP	0.029 (0.022)	0.029 (0.022)	0.029 (0.022)	0.029 (0.022)	0.029 (0.022)
VOLUME	−0.000 (0.000)	−0.000 (0.000)	−0.000 (0.000)	−0.000 (0.000)	−0.000 (0.000)
GENDER	−0.005 (0.011)	−0.005 (0.011)	−0.005 (0.011)	−0.005 (0.011)	−0.006 (0.011)
AGE	0.017 (0.020)	0.016 (0.020)	0.016 (0.020)	0.016 (0.020)	0.016 (0.020)
TENURE	−0.001 (0.002)	−0.001 (0.002)	−0.001 (0.002)	−0.001 (0.002)	−0.001 (0.002)
Constant	0.160 (0.101)	0.143 (0.100)	0.141 (0.100)	0.142 (0.100)	0.141 (0.100)
Firm fixed effects	YES	YES	YES	YES	YES
Year–Quarter fixed effects	YES	YES	YES	YES	YES
R ²	0.1398	0.1405	0.1409	0.1412	0.1415
Adj R ²	0.0480	0.0487	0.0489	0.0494	0.0495
Within R ²	0.0353	0.0361	0.0365	0.0368	0.0371
N	10,787	10,787	10,787	10,787	10,787

Note: Robust standard errors are in parentheses; ***, $p < 0.01$; **, $p < 0.05$; *, $p < 0.1$ (two-tailed).

Data availability

Data will be made available on request.

References

- Ahmed, Y., Elsayed, M., & Xu, B. (2024). Bankruptcy in the UK: Do managers talk the talk before walking the walk? *Br. J. Manag.*, 35(4), 2011–2031.
- Alam, N., & So, J. (2020). Contributions of emotional flow in narrative persuasion: An empirical test of the emotional flow framework. *Commun. Q.*, 68(2), 161–182.
- Amiriparian, S., Gerczuk, M., Lutz, J., Strube, W., Papazova, I., Hasan, A., Kathan, A., & Schuller, B. W. (2024, November). *Non-invasive suicide risk prediction through speech analysis*. 2024 E-Health and Bioengineering Conference (EHB), Romania.
- Amoozegar, A., Berger, D., Cao, X., & Pukthuanthong, K. (2020). Earnings conference calls and institutional monitoring: Evidence from textual analysis. *J. Financ. Res.*, 43(1), 5–36.
- An, K., Chen, Q., Deng, C., Du, Z., Gao, C., Gao, Z., Gu, Y., He, T., Hu, H., & Hu, K. (2024). Funaudiollm: Voice understanding and generation foundation models for natural interaction between humans and llms. arXiv preprint arXiv:2407.04051.
- Aurier, P., & Guintcheva, G. (2015). The dynamics of emotions in movie consumption: A spectator-centred approach. *Int. J. Arts Manag.*, 17(2), 5–18.
- Baker, A. C., Larcker, D. F., & Wang, C. C. (2022). How much should we trust staggered difference-in-differences estimates? *J. Financ. Econ.*, 144(2), 370–395.
- Berger, J., Kim, Y. D., & Meyer, R. (2021). What makes content engaging? How emotional dynamics shape success. *J. Consum. Res.*, 48(2), 235–250.
- Bergh, D. D., Ketchen, D. J., Jr, Orlandi, I., Heugens, P. P., & Boyd, B. K. (2019). Information asymmetry in management research: Past accomplishments and future opportunities. *J. Manag.*, 45(1), 122–158.
- Berkman, H., & Truong, C. (2009). Event day 0? After-hours earnings announcements. *J. Account. Res.*, 47(1), 71–103.
- Blankespoor, E., deHaan, E., & Marinovic, I. (2020). Disclosure processing costs, investors' information choice, and equity market outcomes: A review. *Journal of Accounting and Economics*, 70(2–3), Article 101344.
- Bleiberg, J. (2021). Does the Common Core have a common effect? An exploration of effects on academically vulnerable students. *AERA Open*, 7, Article 23328584211010727.
- Boersma, P., & Van Heuven, V. (2001). Speak and unSpeak with PRAAT. *Glot International*, 5(9/10), 341–347.
- Call, A. C., Flam, R. W., Lee, J. A., & Sharp, N. Y. (2024). Managers' use of humor on public earnings conference calls. *Rev. Acc. Stud.*, 29(3), 2650–2687.
- Cengiz, D., Dube, A., Lindner, A., & Zipperer, B. (2019). The effect of minimum wages on low-wage jobs. *Q. J. Econ.*, 134(3), 1405–1454.
- Chakraborty, B., & Bhattacharjee, T. (2020). A review on textual analysis of corporate disclosure according to the evolution of different automated methods. *J. Financial Rep. Account.*, 18(4), 757–777.
- Chen, Y., Han, D., & Zhou, X. (2023). Mining the emotional information in the audio of earnings conference calls: A deep learning approach for sentiment analysis of securities analysts' follow-up behavior. *Int. Rev. Financ. Anal.*, 88, Article 102704.

- Chou, C.-C., Chang, C. J., Chin, C.-L., & Chiang, W.-T. (2018). Measuring the consistency of quantitative and qualitative information in financial reports: A design science approach. *J. Emerging Technol. Account.*, 15(2), 93–109.
- Chung, S., Cheon, J., & Lee, K.-W. (2015). Emotion and multimedia learning: An investigation of the effects of valence and arousal on different modalities in an instructional animation. *Instr. Sci.*, 43, 545–559.
- Chung, S., & Lee, S. Y. (2019). Visual CSR messages and the effects of emotional valence and arousal on perceived CSR motives, attitude, and behavioral intentions. *Commun. Res.*, 46(7), 926–947.
- Chung, S., & Lee, S. Y. (2020). Cognitive processing of corporate social responsibility campaign messages: The effects of emotional visuals on memory. *Media Psychol.*, 23(2), 244–268.
- Chung, S., & Sparks, J. V. (2016). Motivated processing of peripheral advertising information in video games. *Commun. Res.*, 43(4), 518–541.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Routledge.
- Canyon, M. J., Hass, L. H., Peck, S. I., Sadler, G. V., & Zhang, Z. (2019). Do compensation consultants drive up CEO pay? Evidence from UK public firms. *Br. J. Manag.*, 30(1), 10–29.
- De Amicis, C., Falconieri, S., & Tastan, M. (2021). Sentiment analysis and gender differences in earnings conference calls. *Finance*, 71, Article 101809.
- De Chaisemartin, C., & d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *Am. Econ. Rev.*, 110(9), 2964–2996.
- DellaVigna, S., & Pollet, J. M. (2009). Investor inattention and Friday earnings announcements. *J. Financ.*, 64(2), 709–749.
- Dietrich, B. J., Enos, R. D., & Sen, M. (2019). Emotional arousal predicts voting on the US supreme court. *Polit. Anal.*, 27(2), 237–243.
- DiRusso, C., & Myrick, J. G. (2021). Sustainability in CSR messages on social media: How emotional framing and efficacy affect emotional response, memory and persuasion. *Environ. Commun.*, 15(8), 1045–1060.
- Doey, B., & de Jong, P. (2025). How negative tones in earnings calls shape media narratives. *Rev. Behav. Finance*, 17(3), 406–423.
- Dong, J. Q., & Wu, W. (2015). Business value of social media technologies: Evidence from online user innovation communities. *J. Strateg. Inf. Syst.*, 24(2), 113–127.
- Elliott, W. B., Loftus, S., & Winn, A. (2024). To read or to listen? Does disclosure delivery mode impact investors' reactions to managers' tone language? *Contemp. Account. Res.*, 41(1), 7–38.
- Eriksson, A., & Lacerda, F. (2007). Charlatanry in forensic speech science: A problem to be taken seriously. *Int. J. Speech, Language and the Law*, 14(2), 169–193.
- Ewertz, J., Knickrehm, C., Nienhaus, M., & Reichmann, D. (2025). Listen closely: Using vocal cues to predict future earnings. <https://doi.org/10.2139/ssrn.4307178>.
- Fang, X., Girardone, C., Li, Y., & Zeng, Y. (2025). Generalist CEOs and stock price crash risk. *J. Bus. Financ. Acc.*, 52(1), 182–221.
- Frankel, R., Jennings, J., & Lee, J. (2022). Disclosure sentiment: machine learning vs. dictionary methods. *Manag. Sci.*, 68(7), 5514–5532.
- Gargano, A., & Rossi, A. G. (2018). Does it pay to pay attention? *Rev. Financ. Stud.*, 31(12), 4595–4649.
- Guido, G., Pichierri, M., & Pino, G. (2018). Place the good after the bad: Effects of emotional shifts on consumer memory. *Mark. Lett.*, 29, 49–60.
- Hajek, P. (2022, June). *Speech emotion recognition from earnings conference calls in predicting corporate financial distress*. IFIP International Conference on Artificial Intelligence Applications and Innovations, Berlin.
- Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. The Guilford Press.
- Hutto, C., & Gilbert, E. (2014, June). *Vader: A parsimonious rule-based model for sentiment analysis of social media text*. Proceedings of the International AAAI Conference on Web and Social Media, Ann Arbor, MI.
- Jiang, L., Yin, D., & Liu, D. (2019). Can joy buy you money? The impact of the strength, duration, and phases of an entrepreneur's peak displayed joy on funding performance. *Acad. Manag. J.*, 62(6), 1848–1871.
- Jin, X., & Yu, J. (2022). Does communication increase investors' trading frequency? evidence from a Chinese social trading platform. *Financ. Innov.*, 8(1), Article 68.
- Keene, J. R., & Lang, A. (2016). Dynamic motivated processing of emotional trajectories in public service announcements. *Commun. Monogr.*, 83(4), 468–485.
- Khandeparkar, K., Motiani, M., & Sharma, A. (2021). Thank you for not smoking—A multi-method investigation to understand the effect of anti-smoking warnings in television programs. *J. Bus. Res.*, 128, 462–472.
- Knight, S., Rocklage, M. D., & Bart, Y. (2024). Narrative reversals and story success. *Sci. Adv.*, 10(34), Article ead12013.
- Kuppens, P., & Verduyn, P. (2017). Emotion dynamics. *Curr. Opin. Psychol.*, 17, 22–26.
- Lacerda, F. (2009). *LVA-technology: The illusion of "lie detection"*, FONETIK 2009. Stockholm University, Department of Linguistics.
- Lang, A. (2000). The limited capacity model of mediated message processing. *J. Commun.*, 50(1), 46–70.
- Lang, A. (2006). Using the limited capacity model of motivated mediated message processing to design effective cancer communication messages. *J. Commun.*, 56(s1), S57–S80.
- Lang, A. (2009). The limited capacity model of motivated mediated message processing. In R. L. Nabi, & M. B. Oliver (Eds.), *The SAGE handbook of media processes and effects* (pp. 193–204). Sage.
- Leone, A. J., Minutti-Meza, M., & Wasley, C. E. (2019). Influential observations and inference in accounting research. *Account. Rev.*, 94(6), 337–364.
- Liu, H., Peng, L., & Tang, Y. (2023). Retail attention, institutional attention. *J. Financ. Quant. Anal.*, 58(3), 1005–1038.
- Loh, R. K. (2010). Investor inattention and the underreaction to stock recommendations. *Financ. Manag.*, 39(3), 1223–1252.
- Lortie, C. L., Deschamps, I., Guitton, M. J., & Tremblay, P. (2018). Age differences in voice evaluation: From auditory-perceptual evaluation to social interactions. *J. Speech Lang. Hear. Res.*, 61(2), 227–245.
- Lyle, M. R., Stephan, A., & Yohn, T. L. (2025). Processing time and the speed of the market response to earnings announcements. <https://doi.org/10.2139/ssrn.3064160>.
- Matsumoto, D., Pronk, M., & Roelofs, E. (2011). What makes conference calls useful? The information content of managers' presentations and analysts' discussion sessions. *Account. Rev.*, 86(4), 1383–1414.
- Mayew, W. J., & Venkatachalam, M. (2012). The power of voice: Managerial affective states and future firm performance. *J. Financ.*, 67(1), 1–43.
- Mayew, W. J., & Venkatachalam, M. (2013). Speech analysis in financial markets. *Foundations and Trends® in Accounting*, 7(2), 73–130.
- Meng, T., Lu, D., Yu, D., Yahya, M., & Zariyawati, M. A. (2024). Is executive compensation aligned with the company's ESG objectives? Evidence from Chinese listed companies based on the PSM-DID approach. *Humanit. Soc. Sci. Commun.*, 11(1), 1–15.
- Mohammad, S. (2018, July). *Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words*. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Australia.
- Momtaaz, P. P. (2021). CEO emotions and firm valuation in initial coin offerings: An artificial emotional intelligence approach. *Strateg. Manag. J.*, 42(3), 558–578.
- Moradi, M., Dass, M., & Kumar, P. (2023). Differential effects of analytical versus emotional rhetorical style on review helpfulness. *J. Bus. Res.*, 154, Article 113361.
- Mysková, R., & Hájek, P. (2020). Mining risk-related sentiment in corporate annual reports and its effect on financial performance. *Technol. Econ. Dev. Econ.*, 26(6), 1422–1443.
- Nabi, R. L., & Green, M. C. (2015). The role of a narrative's emotional flow in promoting persuasive outcomes. *Media Psychol.*, 18(2), 137–162.
- Ning, J., Zhang, L., Mi, B., Yang, J. H., & Tao, L. (2024). Signalling through managerial tone and analysts' response. *Account. Forum*, 1–27.
- O'hannon, J. F. (1981). Boredom: practical consequences and a theory. *Acta Psychol.*, 49(1), 53–82.
- Pantazis, C., & Ucar, E. (2014). Religious holidays, investor distraction, and earnings announcement effects. *J. Bank. Financ.*, 47, 102–117.
- Pollock, T. G., Ragozzino, R., & Blevins, P. D. (2024). Not like the rest of us? How CEO celebrity affects quarterly earnings call language. *J. Manag.*, 50(4), 1198–1229.
- Price, S. M., Doran, J. S., Peterson, D. R., & Bliss, B. A. (2012). Earnings conference calls and stock returns: The incremental informativeness of textual tone. *J. Bank. Financ.*, 36(4), 992–1011.
- Price, S. M., Seiler, M. J., & Shen, J. (2017). Do investors infer vocal cues from CEOs during quarterly REIT conference calls? *J. Real Estate Financ. Econ.*, 54(4), 515–557.
- Rambachan, A., & Roth, J. (2023). A more credible approach to parallel trends. *Rev. Econ. Stud.*, 90(5), 2555–2591.
- Rennekamp, K. M., Sethuraman, M., & Steenhoven, B. A. (2022). Engagement in earnings conference calls. *J. Account. Econ.*, 74(1), Article 101498.
- Rhudy, J. L., Williams, A. E., McCabe, K. M., Russell, J. L., & Maynard, L. J. (2008). Emotional control of nociceptive reactions (ECON): Do affective valence and arousal play a role? *Pain*, 136(3), 250–261.
- Robinson, M. D., Storbeck, J., Meier, B. P., & Kirkeby, B. S. (2004). Watch out! That could be dangerous: Valence-arousal interactions in evaluative processing. *Personality and Social Psychology Bulletin*, 30(11), 1472–1484.
- Rojas, S., Kefalianos, E., & Vogel, A. (2020). How does our voice change as we age? A systematic review and meta-analysis of acoustic and perceptual voice data from healthy adults over 50 years of age. *J. Speech Lang. Hear. Res.*, 63(2), 533–551.
- Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *Am. Stat.*, 39(1), 33–38.
- Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychol. Rev.*, 110(1), 145–172.
- Schmidt, M.-L.-C., Winkler, J. R., Appel, M., & Richter, T. (2023). Emotional shifts, event-congruent emotions, and transportation in narrative persuasion. *Discourse Process.*, 60(7), 502–521.
- Schoofs, L., & Claeys, A.-S. (2021). Communicating sadness: The impact of emotional crisis communication on the organizational post-crisis reputation. *J. Bus. Res.*, 130, 271–282.
- Schuller, B., Batliner, A., Steidl, S., & Seppi, D. (2011). Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge. *Speech Comm.*, 53(9–10), 1062–1087.
- Siebert, L., Böck, R., & Wendemuth, A. (2014). Inter-rater reliability for emotion annotation in human-computer interaction: Comparison and methodological improvements. *Journal on Multimodal User Interfaces*, 8, 17–28.
- Signorello, R., Demolin, D., Bernardoni, N. H., Gerratt, B. R., Zhang, Z., & Kreiman, J. (2020). Vocal fundamental frequency and sound pressure level in charismatic speech: A cross-gender and-language study. *J. Voice*, 34(5), Article 808.
- Spiesberger, A. A., Triantafyllopoulos, A., Kathan, A., Semertzidou, A., Gawrilow, C., Reinelt, T., Rauch, W. A., & Schuller, B. (2024, September). "So... my child..."—How Child ADHD Influences the Way Parents Talk. Proceedings of Interspeech, Greece.
- Sprenger, T. O., Sandner, P. G., Tumasjan, A., & Welpe, I. M. (2014). News or noise? Using Twitter to identify and understand company-specific news flow. *J. Bus. Financ. Acc.*, 41(7–8), 791–830.
- Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *J. Econ.*, 225(2), 175–199.
- Sykora, M., Elayan, S., Hodgkinson, I. R., Jackson, T. W., & West, A. (2022). The power of emotions: Leveraging user generated content for customer experience management. *J. Bus. Res.*, 144, 997–1006.

- Tanghe, J., Wisse, B., & Van Der Flier, H. (2010). The formation of group affect and team effectiveness: The moderating role of identification. *Br. J. Manag.*, 21(2), 340–358.
- Throckmorton, C. S., Mayew, W. J., Venkatachalam, M., & Collins, L. M. (2015). Financial fraud detection using vocal, linguistic and financial cues. *Decis. Support Syst.*, 74, 78–87.
- Tiedens, L. Z., & Linton, S. (2001). Judgment under emotional certainty and uncertainty: The effects of specific emotions on information processing. *J. Pers. Soc. Psychol.*, 81(6), 973–988.
- Uparna, J., & Bingham, C. (2022). Breaking “Bad”: Negativity’s benefit for entrepreneurial funding. *J. Bus. Res.*, 139, 1353–1365.
- Wagner, J., Triantafyllopoulos, A., Wierstorf, H., Schmitt, M., Burkhardt, F., Eyben, F., & Schuller, B. W. (2023). Dawn of the transformer era in speech emotion recognition: Closing the valence gap. *IEEE Trans. Pattern Anal. Machine Intelligence*, 45(9), 10745–10759.
- Wang, Z., & Lang, A. (2012). Reconceptualizing excitation transfer as motivational activation changes and a test of the television program context effects. *Media Psychol.*, 15(1), 68–92.
- Yang, Y., Qin, Y., Fan, Y., & Zhang, Z. (2023). Unlocking the power of voice for financial risk prediction: A theory-driven deep learning design approach. *MIS Q.*, 47(1), 63–96.
- Yin, D., Bond, S. D., & Zhang, H. (2017). Keep your cool or let it out: Nonlinear effects of expressed arousal on perceptions of consumer reviews. *J. Marketing Res.*, 54(3), 447–463.
- Yiu, D. W., Wan, W. P., Chen, K. X., & Tian, X. (2024). A theory of host country sentiments: An illustration in cross-border acquisitions. *Acad. Manag. J.*, 67(4), 1024–1054.
- Yu, J., Lin, J. C., & Tian, F. (2026). Market reactions to old innovation news: Evidence from patent disclosures in China. *Int. Rev. Financ.*, 26(2), Article e70078.
- Yu, S., & Leistikow, D. (2011). Abnormal stock returns, for the event firm and its rivals, following the event firm’s large one-day stock price drop. *Manag. Financ.*, 37(2), 151–172.
- Zhang, M., Xie, K., Cai, S., & Wang, Y. (2025). How sentiment volatility of influencer product recommendation posts affects customer engagement. *J. Bus. Res.*, 200, Article 115597.
- Zhang, S., Li, Y., He, Y., & Liang, R. (2024). Do vocal cues matter in information disclosure? evidence from IPO online roadshows in the SSE STAR market. *Int. Rev. Financ. Anal.*, 93, Article 103229.
- Baojun Ma** is a professor at Business School, Hohai University, who has received his Ph.D. degree from Tsinghua University. His research interests focus on brain-machine intelligence for information behavior, mobile big data analytics, and financial technology.
- Lingyun Zhou** is a lecturer at Shanxi Finance and Taxation College, who has received his Ph.D. degree from Shanghai International Studies University. His research interests include financial technology, information management, and artificial intelligence.
- Yao Mu** is an assistant professor at School of Business and Management, Shanghai International Studies University, who has received her Ph.D. degree from Tsinghua University. Her research interests focus on big data analytics, business intelligence, and explainable AI in management.
- Yi Chen** is an assistant professor at School of Business and Management, Shanghai International Studies University, who has received his Ph.D. degree from Nanyang Technological University. His research interests include online community, online platforms, and social networks.
- Jian Zhang** is a professor at School of Management and Business, Shanghai International Studies University, who has received his Ph.D. degree from Temple University. His research interests focus on empirical corporate finance.