Voice Analytics Applications and Corporate Communication: Current State and Future Research Directions

Conference Paper · April 2024

Tim Scheerschmidt
Universität Augsburg

VOICE ANALYTICS APPLICATIONS AND CORPORATE COMMUNICATION: CURRENT STATE AND FUTURE RESEARCH DIRECTIONS

Completed Research Paper

Tim Scheerschmidt, University of Augsburg, Augsburg, Germany, tim.scheerschmidt@uni-a.de

Dennis Renee Metzler, University of Augsburg, Augsburg, Germany, dennis.metzler@uni-a.de

Abstract

Existing studies show that unstructured data, such as voice, has emerged as the future of business analytics. However, research on voice analytics in the context of corporate communication is relatively new and still in its infancy. Therefore, in this paper, we investigate the current state of research on voice analytics in the context of corporate communication and derive promising future research directions. Our analysis shows that existing articles in this field can be mainly divided into two different application areas: (I) Financial Management and (II) Financial Anomaly Prediction. Based on the results of these application areas and additional articles focusing on methodological foundations, we derived a research framework on voice analytics in the context of corporate communication. Additionally, we present several promising future research directions for each application area.

Keywords: Voice Analytics, Unstructured Data, Corporate Communication, Literature Review.

1 Introduction

Communication is ubiquitous in today's business life (Zerfass, 2008). The way in which people speak about economic means often discloses more about their personality and emotional state than most of us expect (Hildebrand et al., 2020). The human voice can be quantitatively decomposed into a set of vocal features. Especially vocal cues such as intonation, speed, volume, and inflection play a significant role in interpreting a message or generating information (Mayew & Venkatachalam, 2012). Although there is a growing body of research on the role of verbal communication through textual analysis, the study of vocal tones in communication is still in its early stages (Mayew & Venkatachalam, 2013). The relatively new research area dealing with this is referred to as voice analytics. It is roughly concerned with the computer-aided processing of (digitized) audio signals for the extraction of information (Mishra & Sharma, 2016). Overall, voice analytics forms an innovative and promising analytical method to gain additional insights from unstructured data (Mayew and Venkatachalam, 2012).

Of particular interest in this context is the vocal communication, that takes place in organizational contexts, which we call corporate communication (Mayew & Venkatachalam, 2013). Relevant company-related information is communicated through various means, such as board meetings, interviews, and earnings conference calls (Mayew & Venkatachalam, 2013; Yang et al., 2023). Researchers and practitioners in the field of corporate communication are particularly interested in studying what is exactly communicated to stakeholders (Aureli, 2017). However, most studies analyzing the information of corporate communication are only focusing on textual analyses (e.g., Li, 2010). Research studying additional vocal cues in corporate communication is even scarer, even though researchers found that analyzing vocal cues in corporate communication provides various additional information.

Given that research at the intersection of voice analytics and corporate communication is still in its infancy but increasingly promising, it is particularly interesting to investigate what the literature has provided so far and what is still underinvestigated (Yang et al., 2023). For this reason, we propose the following research questions:

RQ1: *What is the current state of research concerning voice analytics applications in corporate communication?*

RQ2: *What are worthwhile future research directions on voice analytics in corporate communication?*

To answer these research questions, we conducted a systematic literature following the guidelines of Webster and Watson (2002). Thereby literature is retrieved across various relevant databases with the help of a defined searchstring. Subsequently, the literature is filtered according to specified exclusion and inclusion criteria, allowing the relevant articles to be identified and then categorized.

The results of this paper are reflected in the development of a framework into which the identified literature can be categorized. The basis of the framework is literature that creates a methodological foundation. Besides this, the articles can basically be divided into two application areas. Application area I is labelled as Financial Management. Here, the articles are broken down into corporate firm performance on the one hand and financial risk and volatility on the other. Beside it the application area II is described under the term financial anomaly prediction. This is divided into financial distress prediction and fraud detection. Complementary to the framework, concrete research questions are elaborated for the respective subcategories. These are of particular interest as the research is still in its nascent stages and at the same time offers enormous potential (Mayew & Venkatachalam, 2013).

In the following, the theoretical foundation of this paper will be provided. Here, both components corporate communication and voice analytics are illuminated. Then, the procedure respectively the methodology for answering the research question is described. In the chapter results, the categorized literature is presented, and future research questions are elaborated. This work closes with a discussion and the conclusion.

2 Theoretical Foundation

2.1 Corporate communication

Communication is ubiquitous in today’s business environment (Zerfass, 2008). Company executives utilize various means, such as annual reports, SEC filings, board meetings, sessions, interviews, press conferences and earnings conference calls with analysts and investors, to communicate relevant company-related information (Mayew & Venkatachalam, 2013; Yang et al., 2023). Within the context of corporate communication, these activities can be divided into internal and external communication. Whereas external communication is meant to be available for the public, internal communication is much more sensitive and is primarily meant to be for company relatives, such as managers (Zerfass, 2008). Besides, internal communications processes are often not recorded in contrast to external ones and are rather communicated directly.

External corporate communication in general is done to inform external stakeholders concerning past, current, and predicted financial figures, as well as current and planned strategic plans. The primarily goal of external corporate communication is reducing information asymmetries between a firm's management and external stakeholders, such as investors. Companies primarily use external communication tools, such as highly regulated SEC filings and rather less regulated conference calls to communicate (Moker et al., 2020). In the US region, companies must publish quarterly earnings reports (10-Q reports) and yearly earnings reports (10-K reports) the so-called SEC fillings (SEC, 2022). In conjunction with the release of these reports, earnings conference calls are hosted, which are public meetings that management holds with outside investors and analysts to discuss financial results and strategic directions and additionally respond to analysts' questions (Qin & Yang, 2019). Existing research shows that the question-and-answer section of the conference call is the most informative (Matsumoto et al., 2011). According to Brown et al. (2004), companies that hold conference calls reduce information asymmetry, making earnings conference calls a valuable source of information with high potential respectively opportunities (Price et al., 2017; Sawhney et al., 2020b).

Researchers and practitioners are interested in studying what is communicated to stakeholders and, therefore, need appropriate tools for data collection and analysis (Aureli, 2017). In this context, mainly text mining is used in terms of textual analysis of corporate communication tools like annual reports, profit announcements, and other corporate documents (Li, 2010). While traditional methods of data analysis have focused on structured data, voice mining or voice analytics opens an innovative and promising way to gain insights also from unstructured data (Mayew & Venkatachalam, 2012). However, this relatively new approach is still in its infancy (Mayew & Venkatachalam, 2013).

2.2 Voice analytics

Apart from financial statements and accounting figures, managers convey economically significant information through verbal and nonverbal channels to external stakeholders, such as capital market participants. Particularly, the human voice can be quantitatively partitioned into several vocal characteristics, which Hildebrand et al. (2020) roughly breaks down into the following four dimensions: (i) time, (ii) amplitude, (iii) frequency, and (iv) spectral. Over the years, research across various disciplines developed many unique and innovative technologies and processes to carry out such a complicated but extremely important analysis (Mishra & Sharma, 2016). This gave rise to the field of voice analytics, a branch of data science that deals with the computer processing of audio signals to generate information (Mishra & Sharma, 2016).

In the following, the typical process for obtaining information in the context of voice analytics is outlined. First, unstructured audio data is parsed by using voice analysis technologies. As mentioned in the previous paragraph, there are several different approaches, technologies, and processes. A frequently used approach for emotion extraction is for example the convolutional neural network (CNN), which is applicable for emotion recognition in speech signals due to its adaptive feature extraction and classification properties, that are different from classical neural networks (Aslan, 2021; Zang et al., 2016). In consequence, CNN can adaptively extract features to eliminate the dependence on human subjectivity or experience (Zang et al., 2016). Another common technique is the layered voice analysis (LVA), which consists of a series of proprietary signal processing algorithms that extract and combine attributes from the voice to identify, in specific, the cognitive level, the emotional level, the thinking level, and the global stress level (Mayew & Venkatachalam, 2012; Hajek, 2022; Hobson et al., 2012). A further analysis tool is PRAAT, a software that allows quantification of a wide range of acoustic characteristics such as pitch, intensity, jitter, flicker, and excitation patterns (Hajek, 2022). After the data have been pre-processed and structured, information can be obtained by using the corresponding data analysis method. Finally, value can be added and possible recommendations for action can be derived. The schematic process of voice analytics can be traced in Figure 1.

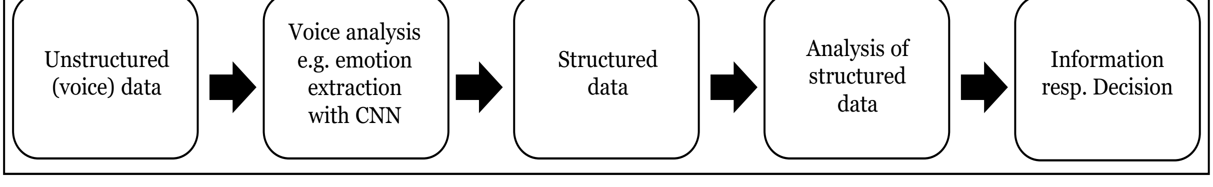

Figure 1. Schematic process of voice analytics

The development and use of voice analytics applications is an entirely new field of research. Research in this area can be found across various disciplines, such as computer sciences, psychology, organizational science, and information systems. Existing studies in this field focus on the development of voice analytic applications (e.g., Qin and Yang, 2019), the analysis of meaning and impact of voice (e.g., Chadwick, 2020; Hiremath and Patil, 2020), voice analytics applications in healthcare (e.g., König et al., 2016), or voice analytics applications in business (e.g., Hildebrand et al., 2020). Voice analytics applications in business are done for various reasons. One research area deals with analyzing corporate communication of companies (e.g., Hajek, 2022). Due to its increasingly high importance for researchers and practitioners alike, in this study, we focus on voice analytics applications in external corporate communications.

3 Methodology

To answer the formulated research questions, we conducted a systematic literature review based on the guidelines of Webster and Watson (2002). This is especially an appropriate research methodology when aiming at analyzing the current state of research and elaborating future research directions concerning a specific research topic. Figure 2 illustrates the different process steps for identifying relevant articles.

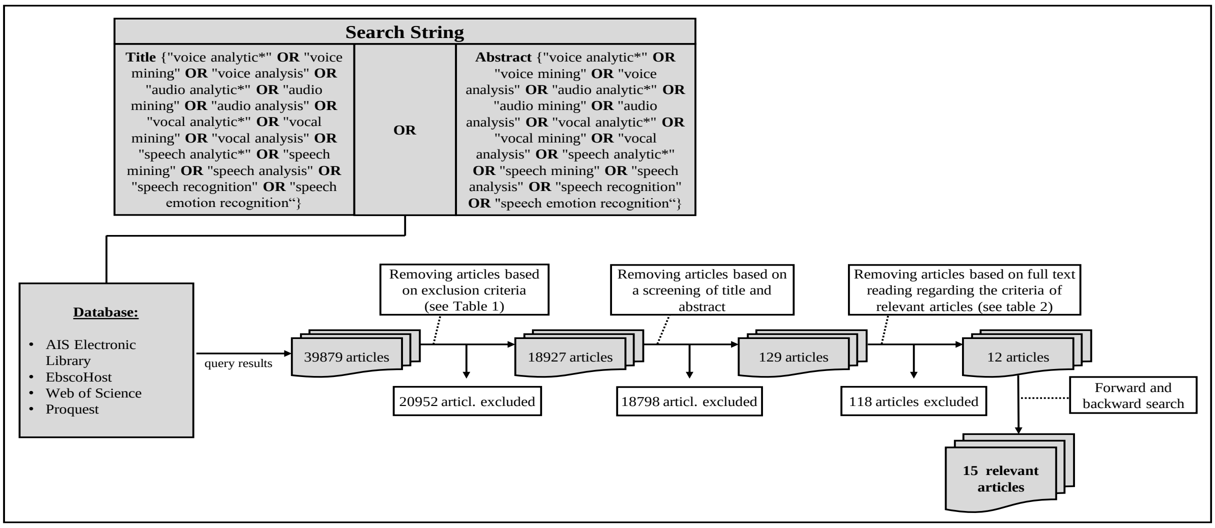

Figure 2. Process of the systematic literature review

The starting point of this literature review was formulating an appropriate search string, based on our research questions. To derive an appropriate search string, we read the basic literature on voice analytics and did a forward- and backward-search to find more appropriate literature. Based on this initial literature search, we derived relevant terms concerning the topic "voice analytics". For example, we found various synonymy, such as "voice recognition", which were added to the search string. Even though this article has the framework "corporate communication", the term is not used in the search string purely for definitional reasons, to include all forms of corporate communication and not to overlook possible articles that only highlight one form of this terminology. In addition, the individual wording and possible communication channels are not included individually in the search string, as this would go beyond the scope of the search and there is also the potential that individual terms could be defined differently or overlooked. Afterward, the derived search string was tested in several databases to ensure its functionality. The final search string was applied to the title and abstract of potentially relevant articles in four different databases. This provides a broad selection of relevant articles. Further, we automatically avoid articles that only partially refer to voice analytics as a small part of the fulltext. The final composition of the search string, as well as the databases chosen, can be obtained from Figure 2. The main identification process of relevant literature was conducted between November 2022 and November 2023. To increase the probability that even the latest publications can be considered in our review, we conducted several queries for each database over time. Thus, we considered the complete publication period till November 2023. After applying the search string on the different databases, duplicates and irrelevant articles were eliminated. Afterward, we eliminated articles based on our exclusion criteria which can be obtained from Table 1.

Criteria	Description
1	Publications not published in a peer-reviewed journal or conference.
2	Articles that are not written in the English language.
3	White papers, commentaries, editorials, and similar articles.

Table 1. Exclusion Criteria

As a next step, we started eliminating articles that are not appropriate to answer our formulated research questions. This primarily concerns articles that do not deal with the topic under investigation. Some articles were eliminated after reading the title and some articles were eliminated after reading the abstract. The process of eliminating articles was conducted by two researchers who independently read through the pre-selected articles. For each article, both researchers decided whether it is relevant or not. After that, the results were compared, and mismatches were discussed to reach a consensus on their relevance. As part of our decision process, we pre-defined some relevance criteria, which can be obtained from Table 2.

Criteria	Description
1	Relevant articles must have a focus on voice analytics applications in business.
2	Relevant articles must have a focus on corporate communication issues.
3	Articles are classified as irrelevant if they only briefly pick up the topic voice analytics or corporate communication.

Table 2. Inclusion Criteria

In our final step of data collection, we conducted a forward and backward search based on the remaining relevant articles. During this last step, we found 3 additional articles not included in the previous database search. This led to a total of 15 relevant articles suitable for answering our research questions.

After selecting the final sample of relevant articles, we analyzed the sample to derive relevant concepts, which finally will be transferred to a concept matrix. We decided to conduct a qualitative content analysis, based on the guidelines of Mayring (2004). We started our content analysis by reading the identified articles and coding all excerpts that refer to voice analytics applications in the context of corporate communication. In a second step, we related all identified codes to each other to derive first subcategories. In this step we identified the subcategories “methodological foundations”, “financial risk and volatility”, “corporate firm performance”, “financial distress prediction”, and “fraud detection”. Finally, we related all subcategories to identify our main categories. The subcategory “methodological foundations” was directly reformed to a main category. The four remaining subcategories were assigned to the main categories “financial management” and “financial anomaly detection”. The emerged subcategories and main categories represent the relevant concepts of our literature review. In a final step, the relevant concepts were finally transferred to a concept matrix (Webster and Watson, 2002). To ensure reliability, the coding process has been always done by two researchers.

4 Results

The principal outcomes of our literature review are exemplified through a concept matrix that links the identified research articles with the designated focus areas (Webster and Watson 2002). The conclusive concept matrix can be acquired from Table 3.

Authors & Year	Methods	Found-ations	Financial Management		Financial Anomaly Detection	
			Firm Performance	Financial Risk	Distress Prediction	Fraud Detection
Burgoon et al. (2016)	PRAAT, MANOVA					X
Hajek (2022)	SER, LSTM				X	
Hildebrand et al. (2020)	none	X				
Hobson et al. (2012)	LVA, MLR					X
Mayew & Venkatachalam (2012)	LVA, MLR		X			
Mayew & Venkatachalam (2013)	LVA, MLR		X			
Mayew et al. (2020)	PRAAT, MLR			X		
Price et al. (2017)	LVA, MLR		X	X		
Qin & Yang (2019)	PRAAT, MDRM			X		
Sawhney et al. (2020a)	PRAAT, DNN			X		
Sawhney et al. (2020b)	PRAAT, LSTM			X		
Sawhney et al. (2020c)	PRAAT, SVR			X		
Throckmorton et al. (2015)	PRRAT, GLRT					X
Yang et al. (2020)	PRAAT, HTML			X		
Yang et al. (2023)	PRAAT, LSTM			X		
Legend: DNN: Deep neural network; GLRT: Generalized likelihood ratio test; HTML: Hierarchical transformer-based multi-task learning; LSTM: Long short-term memory; MANOVA: Multivariate analysis of variance; MDRM: Multi-modal deep regression model; MLR: Multivariate linear regression; SER: Speech emotion recognition (CNN-based); SVR: Support vector regression.						

Table 3. Results of the systematic literature review

In addition to the conceptual matrix, the results were subjected to a descriptive analysis, shedding light on the evolving landscape of voice analytics in corporate communications, as depicted in Figure 3. This figure showcases the escalating scholarly interest in this area, with research endeavors commencing in 2012 and steadily gaining momentum over the years.

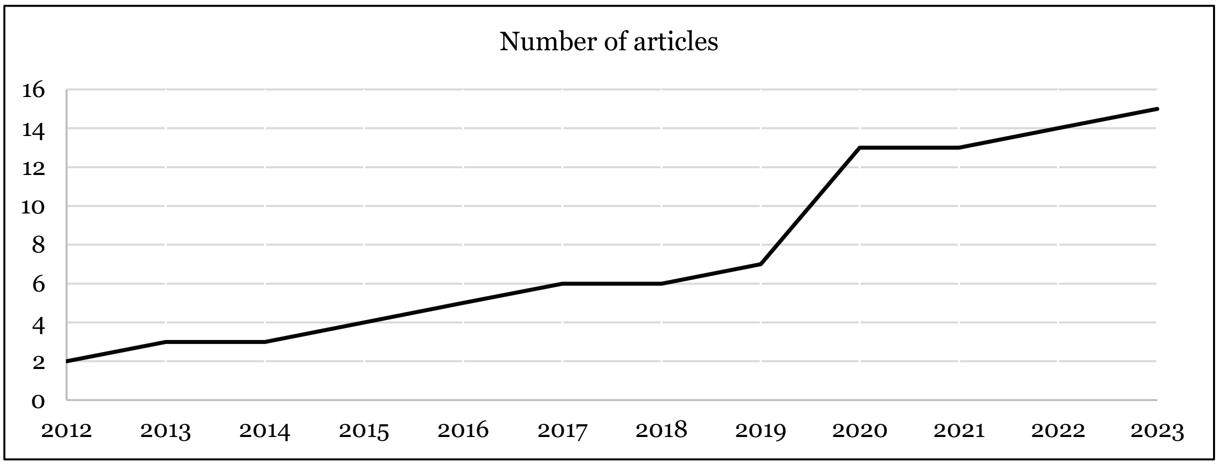

Figure 3. Number of articles per discipline Cumulative numbers of articles per year

Research on voice analytics in the context of corporate communication is done across various scientific disciplines. Most of the articles have been published in the information systems field or finance field, while communication & psychology, business, and computer science have about the same manageable number of publications. The number of articles per subject area is shown in Figure 4.

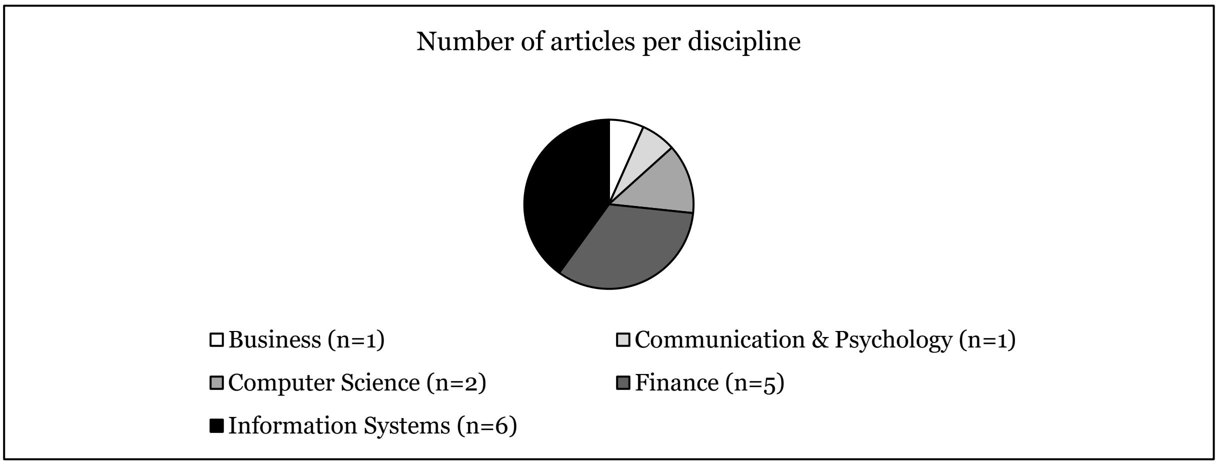

Figure 4. Number of articles per discipline

The existing research studies show different motivations and research questions for the individual research areas. For example, studies from the field of finance tend to neglect aspects such as feature engineering, as Mayew & Venkatachalam (2012) note. On the other hand, research in computer science, for example Qin & Yang (2019), Yang et al. (2020) and Sawhney et al. (2020a, 2020b), focuses primarily on model development to improve prediction accuracy. In contrast, studies on information systems such as Yang et al. (2023) integrate business implications with advanced model design, representing a unique approach.

Based on the results of the systematic literature review and the subsequent analysis of the final articles, the identified literature can be assigned to different categories, which finally form a research framework. This framework on voice analytics applications and corporate communications is illustrated in Figure 5. The basis of the framework is formed by the Methodological Foundation, where fundamental papers are located that take on a superordinate role. Besides this, the articles can basically be divided into two application areas. Application area I is labelled as financial management. Here, the articles are broken down into corporate firm performance on the one hand and financial risk and volatility on the other. Corporate firm performance comprises analyses concerning the financial performance of the company and the associated investor behavior, while financial risk and volatility describes analyses on stock price volatility and earnings predictions as well as financial risks. Application area II is described under the term financial anomaly prediction. This is divided into financial distress prediction, which refers to the prediction of financial difficulties, and fraud detection, which covers the detection of fraudulent actions such as misreporting. In the following, the individual categories are underpinned with the indexed literature.

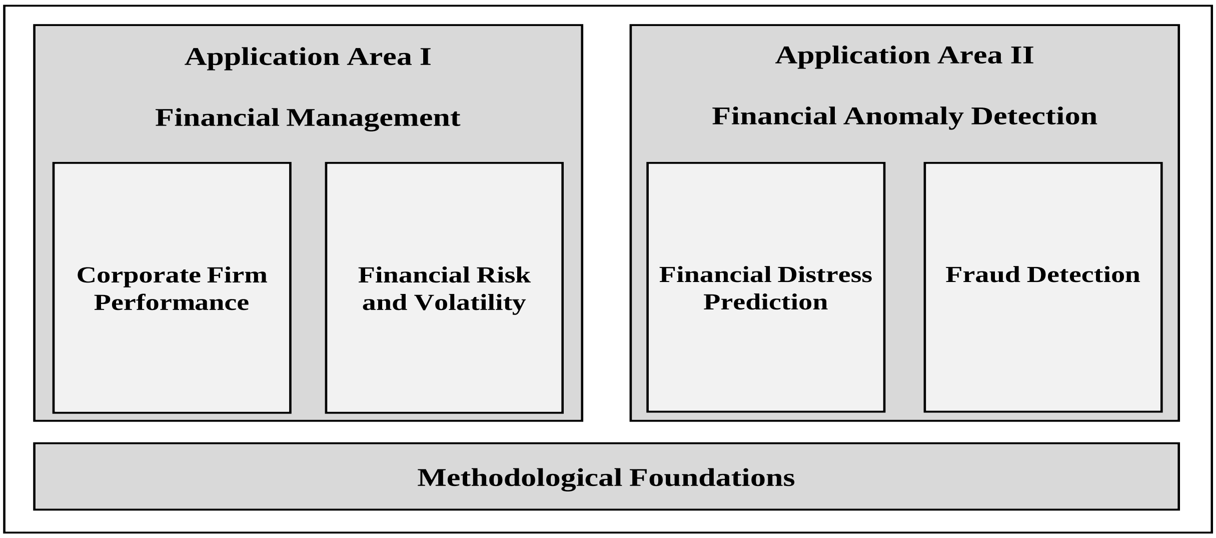

Figure 5. Framework for the analyzed articles

4.1 Methodological foundations

In the methodological basis, the only article assigned to the fundamental works is presented, which plays a general role. In this regard Hildebrand et al. (2020) have developed in the context of business research a four-dimensional integrative conceptual framework that decomposes vocal features of the human voice. The four dimensions of the framework for vocal feature extraction are (i) time, (ii) amplitude, (iii) frequency, and (iv) spectral. The time encompasses duration, speech and articulation rate, and voice breaks, which can reveal information about a speaker's fluency, pacing, and possible vocal fatigue or hesitation. The amplitude dimension focuses on the intensity or power and variability of intensity, which can provide information about a speaker's vocal projection, emotional state, and possible changes in these factors over time. The frequency subcategory includes measures of fundamental frequency, such as pitch and variability in pitch, which can offer insights into a speaker's vocal range, intonation, and potential emotional cues. Finally, the spectral dimension encompasses vocal shimmer, vocal jitter, harmonic to noise ratio, and vocal entropy, which can reveal information about a speaker's vocal stability, consistency, and potential emotional cues (Hildebrand et al., 2020). With this framework, a concept is presented that should support the analysis of speech and serve as an orientation. In addition, the article relates these dimensions and characteristics to various speaker states and traits, which can be seen in detail in the study by Hildebrand et al. (2020).

4.2 Application area I – financial management

Corporate firm performance

This subsection deals with articles relating to the company's financial performance and the associated investor behavior. Mayew & Venkatachalam (2012) utilized vocal emotion analysis software to evaluate the emotional states of managers during earnings conference calls. These affective states represent responses to events and stimuli and can provide valuable insights. The authors discovered that positive and negative emotions displayed by managers during analyst scrutiny on conference calls can indicate the firm's financial profitability and returns. Their study demonstrated that vocal cues from managers can offer crucial information regarding a firm's fundamentals, which is in addition to quantitative earnings data and qualitative "soft" information conveyed by language. The connection remains significant even when controlling for quantitative information and word usage by managers during the call. These results imply that affective states hold an incremental value to linguistic tone in predicting future unexpected earnings, particularly when managers fail to meet analysts' current-period earnings estimates. Investors tend to respond to vocal cues by correlating cumulative abnormal returns with variables that measure the emotional aspects of managers' voices. Based on these findings, it can be concluded that nonverbal vocal cues have a significant influence on future firm performance, particularly when managers are under analysts' scrutiny. More positive (negative) emotions are indicative of a two-quarter-ahead increase (decrease) in future earnings (Mayew & Venkatachalam, 2012). Price et al. (2017) provide the explanation for the temporal shift that call transcripts are presently more widely accessed than call recordings, suggesting that investors reread the language of a call in subsequent days and months, constantly updating their comprehension of the subtle cues it contains. In a subsequent investigation using distinct data yet similar methodology, Mayew & Venkatachalam (2013) corroborated the findings of their prior research in Mayew & Venkatachalam (2012).

Investors search for company-specific information from a wide variety of sources. Conventional forms of corporate disclosure such as SEC filings, press releases, and earnings announcements are closely examined for data that can be plugged into valuation models. Additionally, investors strive for chances to engage with managers on a more individual basis, like analyst conference calls, to gain a better understanding of how they view the firm's prospects. The conversational environment of conference calls provides investors with an opportunity to listen to managers clarify recent performance, offer insights into prospects, and address analyst questions. While comparing expected and realized earnings as reported by managers is likely to provoke an emotional response from investors, assessing how investors react to actual executive emotion is an entirely different matter. The researchers discovered that executive emotion is positively associated with investors' initial reaction, with results that are both statistically and economically significant. Furthermore, this strong investor reaction to emotional cues from managers seems to be justified, indicating that credible, value-relevant information is conveyed by the emotional signals. However, they also observed some limited evidence of a partial reversal in subsequent trading windows, indicating that investors may second-guess themselves or fear they overreacted. Specifically, raw manager emotion is positively and significantly related to cumulative abnormal portfolio returns of roughly 2% shortly after earnings conference calls. Interestingly, they also found some limited evidence of a partial reversion. A visual representation of portfolio returns to high-emotion observations reveals a sharp and rapid initial response to the emotional signals, followed by a subsequent reversal about a week after the call. In addition, the researchers were able to demonstrate that the investors' reaction to the emotional cues from managers was not misguided since the vocal indicators were positively and significantly correlated with the earnings surprise that occurred in the following quarter (Price et al., 2017).

The study of Price et al. (2017) complements the studies of Mayew & Venkatachalam (2012, 2013) by showing that investors also react more quickly to emotionally charged information in the sense of the subsequent quarter and not only two-quarter-ahead.

Financial risk & volatility

Stock price volatility is a measure describing deviations from expected returns and is often used as an indicator of financial risks and thus crucial for investment decision making. (Sawhney et al., 2020a; Qin & Yang, 2019). A research study has presented proof of how the stock recommendations and earnings predictions of individual analysts are connected to the significance of discussions that take place during conference calls between managers and analysts. The authors of the study used the absolute stock price reactions during specific dialogues between managers and analysts to gauge their informativeness. The results showed that dialogs with bearish analysts whose predictions were not met were more informative. Such analysts had longer conversations with more back-and-forth exchanges and exhibited a more negative tone compared to bullish analysts who gave predictions that could be beaten. The direction of stock prices responded to both the linguistic tone of the analyst and the voice pitch of the manager. Therefore, the impact on capital markets during earnings conference calls is more complex than what was previously reported. However, there were no differences in the displays of dominance by managers towards different types of analysts, as captured by voice pitch changes. In other words, dialogs with analysts generate multiple signals that are informative to market participants, but they do not have a systematic directional effect on stock prices. The structure and sentiment of analyst-manager interactions are influenced by the stock recommendation and forecast error of individual analysts, and market participants pay attention to these dialog characteristics. While the study attempted to capture some of the linguistic and acoustic signals conveyed through analyst-manager interactions in a conference call setting, the signals analyzed were not exhaustive or mutually exclusive (Mayew et al., 2020).

Regarding affective states, analysts display asymmetric reactions where a correlation is found between positive affective states and recommendation changes, but not between negative affective states and recommendation changes. Consequently, investors in the stock market tend to underestimate the significance of verbal signals conveying negative affect. Nonetheless, it is not argued that such inadequate response can be employed as a profitable trading approach. This assertion necessitates an extensive evaluation of the effect of both trading expenses and data acquisition costs, which would require a more prolonged historical data set (Mayew & Venkatachalam, 2012). On the other hand, Price et al. (2017) investigate whether raw executive emotion influences revisions in analyst recommendations or forecasts. Regarding the former, some indications suggest that analysts slightly decrease their recommendations after conversing with executives exhibiting high levels of positive emotions. However, regarding the latter, there is no proof that analysts modify their forecasts based on managerial vocal cues (Price et al., 2017).

The findings from Qin & Yang's (2019) study indicate that their model, which incorporates both verbal and vocal characteristics, results in a significant and meaningful decrease in prediction errors. Their study demonstrates that analyzing the language and voice of CEOs during earnings conference calls can help to forecast a company's financial risk, as indicated by the volatility in stock prices in the days following the conference call. Therefore, they suggest a model that combines various features from transcripts and audio recordings of earnings announcements to predict stock volatility after the earnings announcement (Qin & Yang, 2019). The study conducted by Yang et al. (2020) builds upon the research done by Qin & Yang (2019) and brings about significant enhancements in forecasting short and long-term volatility. The researchers introduce a novel hierarchical, transformer, multi-task architecture to leverage text and audio data from quarterly earnings conference calls in order to predict future price volatility. The study includes a thorough evaluation against various baselines, which shows noteworthy advancements in the precision of predictions.

Further studies regarding to forecasting open models that are intended to increase the efficiency and accuracy of analyses and should serve as an orientation or template for future studies. In this context Sawhney et al. (2020a) propose a neural architecture that collectively employs consistency over speech, text, and inter-stock correlations and show that this can elevate the explanatory value. The authors demonstrate that this approach can enhance the explanatory power, thereby exposing disparities between verbal and vocal cues, and subtle variations between textual and auditory components. These discrepancies may go unnoticed in a unimodal model, as illustrated in a case study (Sawhney et al., 2020a). Complementary Sawhney et al. (2020b) use VolTAGE, a further neural architecture, to show that incorporating speech, text and inter-stock correlations can improve volatility forecasting performance. A multitask solution (heterogeneous multi-task approach) is introduced in the study from Sawhney et al. (2020c). The approach employs specialized textual features and audio attentive alignment to model financial risk and price predictions. The central part of the model is a neural attentive alignment model that focuses on interrelationships between vocal and verbal modalities. The multi-task weight sharing ensemble method presented in this study outperforms, among others, text-only models. The authors support their claims with three case studies. Thus, the approach illustrates the potential of using the CEO's verbal and vocal cues during company earnings calls for financial tasks. (Sawhney et al., 2020c). Yang et al. (2023) developed DeepVoice, a novel nonverbal predictive analytics system for predicting financial risk in quarterly earnings conference calls. To study a company's financial risk days after corporate conference calls, a two-stage deep learning model is presented to integrate managers' sequential vocal and verbal cues. It is shown that DeepVoice delivers remarkably lower risk prediction errors and performs significantly more efficiently compared to existing state-of-the-art methods. Moreover, through a cross-sectional analysis, DeepVoice is found to provide greater predictive value for companies with higher variance in financial fundamentals. (Yang et al., 2023)

4.3 Application area II – financial anomaly detection

Financial distress prediction

A study was designed to ascertain the impact of emotions expressed during earnings conference calls on forecasting corporate financial distress. The outcomes of this exploration suggest that state-of-the-art speech emotion recognition technology offers significant insights for financial difficulty forecasting models. More specifically, the emotion happiness signaled a good financial condition, while the emotion sadness indicated financial difficulties. In trend prediction, however, the effect of the emotions was ambiguous. Regardless, the emotions calm and happiness suggested an improvement of the financial situation, while the emotions anger and sadness rather indicated a worsening. The discoveries provide proof of the influence of executives' emotional conditions in identifying corporate financial distress at an early stage. Moreover, it is demonstrated that the proposed deep learning-based prediction model outperforms state-of-the-art financial distress prediction models based solely on financial indicators (Hajek, 2022).

Fraud detection

This category focuses on literature dealing with the detection of fraudulent actions such as misreporting or restatement. Hobson et al. (2012) provide evidence that crucial nonverbal cues are present in earnings conference calls that can be used to detect misreporting. It has been demonstrated through empirical analysis that markers of vocal incongruity can be effective in identifying financial misreporting. In a controlled laboratory environment, a speech sample consisting of both misreporters and truth tellers is produced to establish the validity of the vocal dissonance marker. The findings reveal that vocal dissonance markers generated at the outset of the speech samples, precisely when the dissonance is expected to be most pronounced, are positively linked with four measures of dissonance resulting from misreporting, providing further validation for the cognitive dissonance measure. In an archival setting, CEO speeches that exhibit cognitive dissonance can anticipate if a firm's quarterly financial reports will be restated negatively, surpassing chance levels. The predictive power of vocal dissonance markers is complementary to accounting-based predictors of adverse irregularity restatements (Hobson et al., 2012).

Moreover, disparities in linguistic and vocalic mannerisms have been discovered between the scripted declarations and the question-and-answer segments of conference conversations. Phrases that pertain to the topic of restatement exhibit greater vocal pitch and inferior vocal excellence than unrelated statements. Likewise, expressions regarding restatement topics are spoken at a swifter pace compared to unrelated statements (Burgoon et al., 2016).

A study analyzed the possibility of creating a better tool for detecting financial fraud by merging particular aspects obtained from financial data and the speech of CEOs, as opposed to studying those aspects separately. The findings of this paper endorse the theory that various feature groups including numerical financial information, language patterns, and non-verbal cues from speech could contribute supplementary data for identifying financial fraud. The study found that combining features from different categories resulted in more effective fraud detection compared to relying solely on a single feature category (Throckmorton et al., 2015).

4.4 Future research directions

Concepts	Exemplary Research Questions
Methodological Foundations	• How can voice analytics methods be conceptualized? • What is the most effective method or model for voice analytics?
Corporate Firm Performance	• To what extent can voice analytics be used to measure employee productivity? • Can voice analytics accurately predict revenue growth?
Financial Risk and Volatility	• What are the key features of voice data that can be used to measure financial risk? • What is the accuracy of using voice analytics to predict financial risk and volatility compared to traditional financial models? • Can voice analytics be integrated with other financial analysis tools to improve risk management strategies? • What are the limitations of using voice analytics to predict financial risk and volatility in different market conditions? • How reliable are voice analytics models in predicting financial risk and volatility in emerging markets?
Financial Distress Prediction	• How do the results on financial distress prediction compare to other methods? • What is the most effective method of integrating voice analytics into existing financial distress prediction models? • How accurate are voice analytics predictions compared to traditional financial analysis techniques? • To what extent can voice analytics predict financial distress in small and medium-sized enterprises? • Can voice analytics be used to predict financial bubbles and market crashes?
Fraud Detection	• What specific vocal characteristics or features are most predictive of fraudulent behavior in financial context? • How can voice analytics be integrated into existing fraud detection systems to improve accuracy and efficiency? • What is the accuracy of voice analytics in detecting financial fraud compared to other fraud detection methods? • What are the challenges associated with implementing voice analytics for financial fraud detection in real-world settings? • Can voice analytics be used to identify patterns of deception and manipulation? • Can voice analytics be used to detect and prevent insider trading?
Additional Research Opportunities	• How does the use of voice analytics impact the privacy of individuals and what measures can be taken to address these concerns? • What are the ethical implications of using voice analytics? • How does the accuracy of voice analytics vary across languages? • How do factors such as age, gender, and culture influence the accuracy of voice analytics? • What are the challenges associated with implementing voice analytics? • How can voice analytics be used in conjunction with other data sources to develop more comprehensive models? • How can theory-based methods for extracting language features be developed? • How can established interdisciplinary theories be effectively translated into practical algorithms for voice feature extraction? • What are the challenges in acquiring reliable multimodal datasets in different contexts and how can these challenges be overcome? • How can existing literature be taxonomized based on the challenges encountered in the implementation of voice analytics? • What are the theoretically based and empirically validated dimensions and characteristics for describing and classifying voice analytics use cases?

Table 4. Exemplary Future Research Directions

Having categorized and highlighted the current state of research on voice analytics applications in corporate communications. This chapter is dedicated to the second research question and the issue of what worthwhile future research directions on voice analytics in corporate communication are. Unstructured data has emerged as the future of finance, where voice analytics is an innovative and promising way to gain insights from unstructured data (Mayew & Venkatachalam, 2012), In addition, present the advances in computerized voice analysis coupled with the increasing availability of audio files promising research opportunities (Mayew & Venkatachalam, 2013). To address this, concrete research questions are worked out, which can be obtained from Table 4.

For the category methodological foundations, the questions arise how voice analytics methods can be conceptualized and what is the most effective method or model for voice analytics. In this context, the research community is encouraged to further research and possibly agree on uniform or the most effective methods.

In terms of corporate financial performance, we derived research questions like to what extent voice analytics can be used to measure employee productivity and revenue growth. This could help companies to better plan for the financial future. In the area of financial risk and volatility, there are exciting questions about the accuracy of voice analytics for predicting financial risk and volatility compared to conventional financial models or the interesting question of how and whether voice analytics can be combined with other financial analysis tools to improve risk management strategies. For predicting financial distress, questions arise as to whether voice analytics can be used to identify financial distress in small and medium-sized enterprises or even market collapses and potential financial bubbles. This could contribute to the early detection of economic crises and help to counteract them. Finally, the questions of how voice analytics can be integrated into existing fraud detection systems to improve accuracy and efficiency, or what challenges are associated with implementing voice analytics.

Away from the identified categories, there is also a need for further research. For example, questions of privacy and ethical considerations are of utmost relevance with regard to voice analytics. For instance, Hildebrand et al. (2020) call for an improvement in ethical business practices and analyzing demographic, cultural, and gender biases in context of voice analytics is also a future research direction (Sawhney et al., 2020b).

Corporate earnings announcements are a rich, under-researched source of multimodal information (Sawhney et al., 2020b). In addition, unstructured data, e.g., from social media, may also have high information value (Yang et al., 2020). In addition, various sources of multimedia data, such as press conferences and manager interviews, could also be drawn upon to investigate the information function of voice in financial markets (Yang et al., 2023). For this reason, how can voice analytics be used in conjunction with other data sources to develop more comprehensive models, is a highly promising research prospect. For example, to generate speech samples for analysis, Hobson et al. (2012) play videos of the conferences and manually isolate only the audio of each participant's responses to the interview questions. In this case, video analytics or specifically exploring facial expressions as another channel of nonverbal communication could be a fruitful avenue for future research (Mayew & Venkatachalam, 2012).

Furthermore, future research could delve into the implementation of theory-based methods for the extraction of voice features and thus improve the effectiveness of voice analytics applications. In addition, research into the effective translation of established interdisciplinary theories into practical algorithms for voice feature extraction represents a promising avenue for investigation. Categorizing the existing literature according to the challenges encountered in implementing voice analytics techniques could add valuable insights into best practices and further potential research gaps. Beyond this, addressing the challenges associated with obtaining reliable multimodal datasets in different contexts could be a focus of future studies to develop strategies to overcome these obstacles. Finally, the development of a taxonomies in the field of voice analytics with a focus on use cases or challenges is intriguing, as it would provide a structured framework for organizing knowledge and facilitating understanding of the diverse applications and obstacles within the field.

5 Discussion

5.1 Implications, limitations, and future research

In this study, we did a comprehensive review of the existing literature on voice analytics applications in the corporate communication of companies. Our findings indicate that this area of research is still at a very early stage with a high potential for future research avenues. The novelty of this research area is especially reflected in a very low number of publications. The high potential for this research area is reflected by a very high number of research gaps and associated potential research directions. Based on or review, we show that the current state of literature can be divided into five different areas. This comprises an area of methodological foundations and two different application areas: (I) Financial Management and (II) Financial Anatomy Detection. Both application areas comprise two sub areas. We organized research on voice analytics with a focus on financial risk and volatility and corporate firm performance under the application area one. Application area two comprises studies on financial distress detection and fraud detection. In each of the identified research areas, we found relevant research gaps that can be addressed by future research. Further, we argue for tackling even more research gaps in completely new research areas. For example, we argue for going even beyond voice analytics up to using video analytics applications to additionally include gestures and mimics into analyses.

Despite the careful design of our research approach, this study is subject to some limitations. Based on the limitations, we see a high potential for future research in each case. First, the methodological approach could be enhanced by additionally more databases or even articles in other languages. The search string could also be adjusted by adding even more voice analytics specific key words and adding corporate communication specific key words. Additionally, the search string could be applied on the fulltext of potentially relevant articles instead of limiting the search to the title and abstract. To increase the quality of the final sample of papers, a paper ranking (e.g., the “Financial Times 50”) could be considered. However, due to the novelty of this research area, we have decided against it. Finally, the research focus could be advanced in many ways. For example, it’s possible to choose a wider focus on finance in general. Overall, each of these potential extensions provide the opportunity for gaining more information. However, some of these potential extensions also entail the risk of losing the research focus.

6 Conclusion

Voice Analytics becomes increasingly important since the tone of a spoken word contains important additional information in contrast to text. In this context, unstructured data has emerged as the future of business, where voice analytics is an innovative and promising way to gain insights from unstructured audio and video data. One important application area of voice analytics is corporate communication. However, existing research on voice analytics applications and corporate communication is still in its infancy. Against this background, we aimed to structure the existing literature in this field and derive future research directions. By conducting a systematic literature review (Webster and Watson, 2002), we found that the insights of the existing literature can be classified into five concepts: (i) methodological foundations, (ii) financial risk and volatility, (iii) corporate firm performance, (iv) financial distress prediction, and (v) fraud detection. Research on voice analytics to analyze corporate communication is still rare and future research is encouraged to close research gaps to make the picture of this increasingly relevant research field clearer.

References

Aureli, S. (2017). “A comparison of content analysis usage and text mining in CSR corporate disclosure,” *International Journal of Digital Accounting Research* 17 (1), 1-32.

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

Brown, S., Hillegeist, S. A., and Lo, K. (2004). "Conference calls and information asymmetry," *Journal of Accounting and Economics* 37 (3), 343-366.

Chadwick, R. (2020). "Methodologies of voice: Towards posthuman voice analytics," *Methods in Psychology* 2 (1), 1-10.

Hajek, P. (2022). "Speech Emotion Recognition from Earnings Conference Calls in Predicting Corporate Financial Distress," *in Proceedings of the IFIP International Conference on Artificial Intelligence Applications and Innovations*, Hersonissos, GRC.

Hiremath, B. N., and Patil, M. M. (2019). "Analysis of Voice Cues in Recognition of Sarcasm," *Recent Advances in Computer Science and Communications* 14 (4), 1256-1270.

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

Hildebrand, C., Efthymiou, F., Busquet, F., Hampton, W. H., Hoffman, Donna L., and Novak, T. P. (2020). "Voice analytics in business research: Conceptual foundations, acoustic feature extraction, and applications," *Journal of Business Research* 121 (1), 364-374.

König, A., Satt, A., David, R., and Robert, P. (2016). "Innovative Voice Analytics for the Assessment and Monitoring of Cognitive Decline in People with Dementia and Mild Cognitive Impairment," *in Proceedings of the Alzheimer's Association International Conference*, Toronto, CAN.

Li, F. (2010). "Textual analysis of corporate disclosures: A survey of the literature," *Journal of Accounting Literature* 29 (1), 143-165.

Matsumoto, D., Pronk, M., and Roelofsen, E. (2011). "What Makes Conference Calls Useful? The Information Content of Managers' Presentations and Analysts' Discussion Sessions," *The Accounting Review* 86 (4), 1383-1414.

Mayew, W. J., and Venkatachalam, M. (2012). "The Power of Voice: Managerial Affective States and Future Firm Performance," *Journal of Finance* 67 (1), 1-43.

Mayew, W. J., and Venkatachalam, M. (2013). "Speech Analysis in Financial Markets," *Foundation and Trends in Accounting* 7 (2), 73-130.

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

Mayering, P. (2014). *Qualitative content analysis: theoretical foundations, basic procedures and software solution*, Klagenfurt, GER: Lebniz-Institut für Sozialwissenschaften.

Mishra, C., and Sharma, A. M. (2016). "A review paper on voice analytics," *International Journal of Science and Technology Management* 5 (12), 247-257.

Moker, A., Brosi, P., and Welpe, I. M. (2020). "It Depends on the Size: How Firm Strategic Emphasis on Digital Transformation Predicts Market Capitalization," *in Proceedings of the 53rd Hawaii International Conference on System Sciences*, Hawaii, USA.

Price, S. M., Seiler, M. J., and Shen, J. (2017). "Do Investors Infer Vocal Cues from CEOs During Quarterly REIT Conference Calls?," *Journal of Real Estate Finance and Economics* 54 (1), 515-557.

Qin, Y. and Yang, Y. (2019). "What You Say and How You Say It Matters: Predicting Financial Risk Using Verbal and Vocal Cues," *in Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, ITA.

Sawhney, R., Aggarwal, A., Khanna, P., Mathur, P., Jain, T., and Shah, R. R. (2020a). "Risk forecasting from earnings calls acoustics and network correlations," *in Proceedings of the Interspeech*, Shanghai, CHN.

Sawhney, R., Khanna, P., Aggarwal, A., Jain, T., Mathur, P., and Shah, R. R. (2020b). "VolTAGE: Volatility Forecasting via Text Audio Fusion with Graph Convolution Networks for Earnings Calls," *in Proceedings of the Conference on Empirical Methods in Natural Language Processing*, online.

Sawhney, R., Mathur, P., Mangal, A., Khanna, P., Shah, R. R., and Zimmermann, R. (2020c). "Multi-modal multi-task financial risk forecasting," *in Proceedings of the 28th ACM International Conference on Multimedia*, Seattle, USA.

SEC. (2023). *Securities Laws, U.S. Securities and Exchange Commission*. URL: https://www.sec.gov/about/ about-securities-laws (visited on July 12, 2023).

Throckmorton, C. S., Mayew, W. J., Venkatachalam, M., and Collins L. M. (2015). "Financial fraud detection using vocal, linguistic and financial cues," *Decision Support Systems* 74 (1), 78-87.

Webster, J., and Watson, R. T. (2002). "Analyzing the Past to Prepare for the Future: Writing a Literature Review," *MIS Quarterly* 26 (2), xiii-xxiii.

Yang, L., Ng, T. L. J., Smyth, B., and Dong, R. (2020). "HTML: Hierarchical Transformer-based Multi-task Learning for Volatility Prediction," *in Proceedings of the World Wide Web Conference*, Taipei TWN.

Yang, Y., Qin, Y., Fan, Y., and Zhang, Z. (2023). "Unlocking the Power of Voice for Financial Risk Prediction: A Theory-Driven Deep Learning Design Approach," *MIS Quarterly* 47 (1), 63-96.

Zerfass, A. (2008). "Corporate Communication Revisited: Integrating Business Strategy and Strategic Communication," in Public Relations Research, Zerfass, A., van Ruler, B., Sriramesh, K. (eds). Wiesbaden, GER: VS Verlag für Sozialwissenschaften.

Zhang, B., Quan, C., and Ren, F. (2016). "Study on CNN in the recognition of emotion in audio and images," *in Proceedings of the IEEE/ACIS 15th International Conference on Computer and Information Science*, Okayama, JPN.