Crawford School of Public Policy

CAMA

Centre for Applied Macroeconomic Analysis

Emotion in Euro Area Monetary Policy Communication and Bond Yields: The Draghi Era

CAMA Working Paper 75/2022
December 2022 (Revised in April 2025)

Dimitrios Kanelis
Deutsche Bundesbank

Pierre L. Siklos
Wilfrid Laurier University
Centre for Applied Macroeconomic Analysis, ANU

Abstract

We combine modern methods from Speech Emotion Recognition and Natural Language Processing with high-frequency financial data to precisely analyze how the vocal emotions and language of ECB President Mario Draghi affect the yields and yield spreads of major euro area economies. This novel approach to central bank communication reveals that vocal and verbal emotions significantly impact the yield curve, with effects varying in magnitude and direction. Our results reveal an important asymmetry in yield changes with positive signals raising German, French, and Spanish yields, while negative cues increase Italian yields. Our analysis of bond spreads and equity markets indicates that positive communication influences the risk-free yield component, whereas negative communication affects the risk premium. Additionally, our study contributes by constructing a synchronized dataset for voice and language analysis.

Keywords

artificial intelligence, asset prices, communication, ECB, high-frequency data, speech emotion recognition

JEL Classification

E50, E58, G12, G14

Address for correspondence:

(E) cama.admin@anu.edu.au

ISSN 2206-0332

Emotion in Euro Area Monetary Policy Communication and Bond Yields: The Draghi Era

Dimitrios Kanelis[1], Pierre L. Siklos[2]

April 2025

Abstract

We combine modern methods from Speech Emotion Recognition and Natural Language Processing with high-frequency financial data to precisely analyze how the vocal emotions and language of ECB President Mario Draghi affect the yields and yield spreads of major euro area economies. This novel approach to central bank communication reveals that vocal and verbal emotions significantly impact the yield curve, with effects varying in magnitude and direction. Our results reveal an important asymmetry in yield changes with positive signals raising German, French, and Spanish yields, while negative cues increase Italian yields. Our analysis of bond spreads and equity markets indicates that positive communication influences the risk-free yield component, whereas negative communication affects the risk premium. Additionally, our study contributes by constructing a synchronized dataset for voice and language analysis.

JEL Classification: E50, E58, G12, G14

Keywords: Artificial Intelligence, Asset Prices, Communication, ECB, High-Frequency Data, Speech Emotion Recognition

---

[1]Corresponding Author: Deutsche Bundesbank, Wilhelm-Epstein-Str. 14, 60431 Frankfurt am Main, Germany; E-mail: dimitrios.kanelis@bundesbank.de

[2]Department of Economics, Wilfrid Laurier University, Balsillie School of International Affairs, and the Centre for Applied Macroeconomic Analysis (CAMA)

[‡]We extend our gratitude to Michelle Alexopoulos, Joscha Beckmann, Martin Bohl, Nicole Branger, Michael Ehrmann, Federico Ferrara, Benny Hartwig, Klodiana Istrefi, Bernd Kempa, Dolly King, Malte Knüppel, Rui Liu, Jan Szczygielski, Doron Reichmann, Peter Tillmann, Christian Wagner, and Johannes Zahner for their valuable comments and remarks. We deeply appreciate the opportunity to present this study at the Term Structure Workshop 2024 of the Deutsche Bundesbank, the European Central Bank's seminar *AI in Economics*, the Finance Research Seminar of Kozminski University, and research seminars at the University of Giessen and University of Hagen. Additionally, we thank the participants at the 2025 Society for Nonlinear Dynamics and Econometrics at the 2025 ASSA Meetings, the ECONDAT Fall meeting 2024 at the Bank of Italy, the annual meeting of the Southern Finance Association 2024, the 2024 ESIF Economics and AI+ML Meeting by the Econometric Society, the annual conference of the International Association for Applied Econometrics 2023, and the International Conference for Macroeconomic Analysis and International Finance 2023. We also appreciate the feedback from participants of internal research seminars at the Deutsche Bundesbank and the University of Münster. Special thanks to Niklas Humann for his exceptional research assistance. Part of the research for this paper was conducted while the first author was a visiting scholar at the Balsillie School of International Affairs. He is grateful for their hospitality. This draft is a revised version of the first draft from December 2022. The views expressed in this paper are those of the authors and do not necessarily reflect the opinions of the Deutsche Bundesbank or the Eurosystem.

"[The Press Conference] was originally an obligation, then it became a welcome obligation, and then even a pleasure. [...] Communication has become a tool of monetary policy, so your interaction has been essential in our monetary policy decisions all throughout these eight years."

Mario Draghi to Journalists, 24 October 2019

"[D]ue to this intervention of the activist at the beginning of the press conference I observed that you remained very calm. How do you manage this?"

Journalist to Mario Draghi, 15 April 2015

1. Introduction

When Mario Draghi, former President of the European Central Bank (ECB), spoke, financial markets listened. Psychologists and linguists have long known that emotions impact decision-making. Only recently, however, has the long-recognized role of emotions, and their measurable implications, been investigated by economists.

As the above quote from a journalist illustrates, it is not just the written word listeners consider but the full range of modes of communication. Mehrabian (1971) describes, and others have since shown (see below), that this includes non-verbal forms of communication. For instance, the New York Times explicitly noted Draghi's annoyance due to a journalist's question. This is suggestive of the potential for emotions to matter to readers of financial news.[1]

In part because central bank communication has grown more uniform over time, with policy makers intent on minimizing the likelihood of surprising observers and financial markets. This has led to greater attention to less conventional signals. Vocal cues, especially, provide spontaneous insights that often transcends literal speech. Gorodnichenko et al. (2023) demonstrate how the Federal Reserve chair's vocal emotions can influence asset prices, un-

[1] The NYT writes: *"'Go back and ask yourself, where were you two years ago?' Mr. Draghi said, with a hint of annoyance in his voice."* See: https://www.nytimes.com/2014/10/03/business/international/ecb-leaves-key-interest-rate-unchanged.html (Last Access: 1 September 2022).

derscoring the critical role of these cues in monetary policy communication. Yet, as we shall see, limitations remain in the literature to quantify vocal emotions in this emerging field of research. Our study takes a step in the direction of overcoming some of the constraints other studies have had to contend with.

In the euro area, one can also highlight several examples illustrating how observers detect non-verbal reactions of former ECB president Mario Draghi, like his calm behavior[2] or his annoyed reaction when a journalist refers to his German critics[3]. It is also worth noting that, since the ECB does not reveal much information about the actual debates or the climate during discussions inside the governing council meetings, there is an additional incentive to look for other forms of information.[4] One can only deduce from non-verbal behavior how satisfied the president is with monetary policy decisions, given the economic outlook and the different views within the council about how monetary policy should be conducted (Brunnermeier et al., 2016).

Our study introduces and applies an improved methodology to capture the nuances and influence of vocal emotions on euro area monetary policy. Our focus is on the behavior of interest rates and interest rate spreads since these represent an important barometer of the current and anticipated state of the euro area economy.

We employ modern Speech Emotion Recognition (SER) and Natural Language Processing (NLP) methods, combined with high-frequency financial data, to estimate how the vocal

---

[2]See the article of Insider: https://www.businessinsider.com/who-was-the-protester-who-got-into-the-ecb-and-glitter-bombed-mario-draghi-2015-4 (Last access: August 4, 2022). Also, the article of Bloomberg: https://www.bloomberg.com/news/articles/2017-05-10/draghi-stays-calm-on-stimulus-as-dutch-warn-of-risks-with-tulip#xj4y7vzkg (Last access: August 4, 2022).

[3]See the article of Independent: https://www.independent.ie/business/world/angry-draghi-fights-back-at-german-ecb-critics-34648966.html (Last access: August 4, 2022)

[4]The monetary policy accounts, which the ECB has been publishing since February 2015, provide only an overview. See: https://www.ecb.europa.eu/press/accounts/html/index.en.html (Last Access: August 1, 2022).

emotions and language in Mario Draghi's introductory statement and answers during the Q&A sessions affect bond yields in the four largest euro area economies, namely France, Germany, Italy, and Spain. We consider government bonds because they are highly liquid financial assets and are widely considered a reference point of overall financing conditions in the euro area.

We also create a novel dataset of synchronized audio and textual data for ECB press conferences (May 2012 – October 2019) and resort to an event study approach to examine their effect on yields and yield spreads of different maturities. One advantage of this period is that automated, voice-based trading algorithms were still in their infancy, so market reactions predominantly reflect individual traders' responses rather than systematic algorithmic signals. As a result, Goodhart's Law does not significantly affect our analysis, because the voice-based measure had not yet become a target that market participants could exploit or manipulate.

However, a significant challenge is that Draghi addresses multiple and varied questions in sequence. To ensure that we accurately measure vocal emotions over a wide variety of topics, we exploit an interesting feature in the structure of ECB press conference transcripts. The ECB staff identifies individual answers or, focal points, and organizes the transcripts in such a way that they can be easily identified. Following this structure, we are able to adjust the audio data for each answer and establish synchronicity between voice and words. While Gorodnichenko et al. (2023) isolate vocal tone, we follow communication research (see Filippi et al. (2017)) which shows that market participants process both content and delivery. Consequently, we can assess how vocal emotions and spoken content interact.

We measure vocal emotions using the Fully Convolutional Neural Network (FCN) by García-Ordás et al. (2021), which handles audio files of varying lengths without excessive preprocessing. This analysis applies both to the introductory statement and the Q&A sessions.

Regarding the verbal cues of the introductory statement, we utilize the sentiment indicator by Geiger et al. (2025), who introduced the Monetary Intelligent Language Agent (MILA) which measures economic optimism and pessimism conditional on the macroeconomic context. To analyze language framing in the Q&A sessions, we employ FinBERT, a BERT-based model specifically adapted for economic and finance-related texts (Huang et al., 2023).

Our findings highlight the importance of vocal emotions and language for monetary policy communication. We estimate a significant impact of vocal emotions and language on yields and yield spreads of major euro area economies. For example, we find that for German and French bonds, positive vocal and verbal cues lead to yield increases. Conversely, negative vocal and verbal cues result in increased yields for Italian bonds, while no effects are estimated for the German, French, and Spanish yields. Furthermore, we find that the Q&A session is the main driver of asset prices, whereas the scripted introductory statement has a comparatively smaller impact. The asymmetry we identify is an important one given the economic circumstances Italy faced during the 2012-19 period we investigate. The fallout from the European Sovereign Debt Crisis (ESDC) loomed large and our results highlight the resulting divide this created in government bond yields.

When yield spreads are examined instead, our results reveal how vocal emotions shape investor risk behavior. Negative cues boost the Italian spread, whereas French and Spanish spreads remain unaffected. This pattern implies that positive unscripted communication acts like a conventional monetary impulse, while negative cues alter the risk premium of specific countries. Equity markets echo these results: negative cues lower stock prices, consistent with broader literature on different market responses to good versus bad news. Once again, we are led to the possibility of financial markets reacting differently to positive versus negative news conditional on the state of the economy of the country in question.

Our study makes four contributions to the literature. First, to the best of our knowledge, we are the first to document the ECB President's vocal emotions as a market-relevant factor. Second, we apply deep learning to economic analysis by using the FCN architecture of García-Ordás et al. (2021), advancing the use of audio data in central bank communication and empirical finance. Its capacity to process non-fixed audio lengths is key for varied-response events like Q&A sessions. Third, we extend the small but growing body of work linking behavioral aspects of monetary communication to financial outcomes. Fourth, we create a new dataset of synchronized voice and textual data, plus additional qualitative information, that can be used in further research.

The remainder of the study is structured as follows. Section 2 provides a literature review on monetary policy event studies, its intersection with central bank communication, and the existing literature that considers vocal emotions. Section 3 extensively describes the methodology used to construct the data set and implement the event study regression. Section 4 presents the results. The final section concludes.

2. Event Studies and Central Bank Communication

Event studies represent an influential approach for analyzing the direct causal effects of monetary policy decisions. Using a narrow time window around an event and utilizing high-frequency data (daily or intra-daily), the causality is identified by disentangling and ordering the sequence of events (Ramey, 2016). The literature has grown considerably since the first seminal studies using this identification strategy (e.g., Kuttner (2001); Cochrane and Piazzesi (2002); Rigobon and Sack (2004)). Gürkaynak et al. (2005) use factor analysis with high-frequency federal funds futures (FFF) data and estimate two factors to explain the FFF interest rate movements in real time. The authors analyze how monetary policy announcements drive the second factor during press conferences. Cieslak and Schrimpf (2019) distinguish between monetary and non-monetary news that markets perceive during central bank press conferences and estimate separate effects on the comovement of stocks and interest rates. The authors find that news composition in central bank statements varies considerably, with non-monetary news driving communication between 2008 - 2013 and monetary news after that. Swanson (2021) extends the identification approach of Gürkaynak et al. (2005) to years following the Great Financial Crisis (GFC) and finds that additional factors are necessary to explain real-time movements of asset prices. These are motivated by the unconventional monetary policies of the Fed.

Event studies focusing on the ECB generally exploit the unique structure of monetary policy decision announcements, that is, the time difference between the press release at 13:45 CET and the press conference at 14:30 CET (e.g., Brand et al. (2010)).[5] Altavilla et al. (2019) use factor analysis on real-time Overnight Index Swap (OIS) data and consider three latent factors necessary to explain the variance of changes in OIS during the press conference since the GFC. Analogous to Swanson (2021), they attribute this to the increased importance of unconventional monetary policy instruments such as forward guidance (FG) and quantitative easing (QE).

The second strand of literature relevant to our study considers the effects of central bank communication via machine learning techniques. Early surveys, such as Blinder et al. (2008), updated recently by Blinder et al. (2024), provide foundational insights. Typical approaches to quantifying qualitative central bank communications include dictionary methods (e.g., Loughran and McDonald (2011); Apel and Blix Grimaldi (2014)) or textual indicators such as complexity or similarity (Ferrara and Angino, 2022), applied to textual data from various central bank outputs like introductory statements (e.g., Picault and Renault (2017)), transcripts (e.g., Shapiro and Wilson (2021)), speeches (e.g., Bohl et al. (2023)), press releases (Ehrmann and Talmi (2020)), and social media contributions (Ehrmann and Wabitsch (2022)). Notable findings include Hubert and Labondance (2021) showing how the tone of

[5]Since July 21, 2022, the ECB has changed the times of its announcement.

introductory statements can explain monetary shocks and predict future policy decisions. Parle (2022) and Schmelling and Wagner (2025) find that the framing of central bank press conference statements affects stock prices in real-time.

Since press releases following policy rate decisions change little most of the time (Ehrmann and Talmi, 2020), markets also want to look for unscripted information that provides additional hints about the conduct of monetary policy or the sentiment of the policy-making committee. The emotional attitude of central bankers is one source of unscripted information that may help explain asset price movements. Indeed, as financial research demonstrates, investors observe vocal emotions. In an early approach to analyzing non-verbal communication, Mayew and Venkatachalam (2012) provide evidence that vocal emotions influence investors and that positive (negative) vocal emotions of managers precede positive (negative) news about corporate performance.

In the case of central banking, Gorodnichenko et al. (2023) measure the vocal sentiment of the Fed chair and estimate a significant effect on stock prices in the days following FOMC press conferences. The authors refer to asymmetric information as a potential explanation for the market's interest in non-verbal behavior. This explanation is also consistent with the argument of Mayew and Venkatachalam (2012), who acknowledge that negative private information contradicting one's beliefs results in an uncomfortable state of mind that manifests itself in corresponding non-verbal communication.[6] Studies by Curti and Kazinnik (2023) and Alexopoulos et al. (2024) estimate real-time effects of the chair's facial emotions on stock prices and find them to be important for our understanding of the conduct of monetary policy. In addition, by looking at the pitch of the voice, Alexopoulos et al. (2024) show that the voice is a significant determinant of the high frequency financial market responses

[6] Hobson et al. (2012) show that vocal dissonance of CEOs is positively associated with the likelihood of financial misreporting.

and in the days following the event.[7]

3. Methodology

To study the effect of the emotion of the ECB president on government bond yields, we adopt an event study approach. During the entire presidency of Mario Draghi, the ECB released monetary policy decisions in a press release at 13:45 CET. At 14:30 CET, the press conference begins with the president reading a prepared introductory statement and then provides a Q&A session for journalists. By analyzing median yield changes within a ten-minute interval after the press conference concludes, we estimate an unbiased and direct causal effect (Altavilla et al., 2019).[8] We utilize data from the *press conference windows* of the Euro Area – Monetary Policy Event-Study Database (EA-MPD) provided by Altavilla et al. (2019).[9] Focusing on median yield changes within this interval helps mitigate the impact of high volatility and noise in financial market data with higher frequency. Although the dataset may not capture the immediate effects of brief emotions during the press conference, it allows us to estimate a real-time effect at the press conference level. Additionally, we ensure our estimations are not influenced by new monetary policy decisions, as financial markets are already aware of them following the press release.

To estimate the effects of emotion and language on government bond yields during the press conference, we estimate the following benchmark regression model:

---

[7]Since the first version of our paper was written, a recent analysis by Barry et al. (2025) has also investigated the role of non-verbal communication in euro area monetary policy. The authors focus on facial expressions to augment vocal and verbal signals and analyze the differences between the Draghi and Lagarde presidencies, building on the literature of earlier studies (including ours).

[8]The yield change occurs several minutes after the press conference ends, ensuring no reverse or simultaneous causality bias (Altavilla et al., 2019).

[9]For further details on deriving asset yields, refer to the second section of their paper and the appendix of their study.

$$
\begin{aligned}
y_t = \beta_0 + \beta_1 * Voice_t^{AN} \times Language_t^{AN} + \beta_2 * Voice_t^{AN} + \\
\beta_3 * Language_t^{AN} + \beta_4 * Voice_t^{IS} + \beta_5 * Language_t^{IS} + \\
\sum_{i=6} \beta_i X_{ti} + \epsilon_t
\end{aligned}
\tag{1}
$$

y_t represents the change in yield for German bonds of one, two, five, or ten-year maturities, or for French, Italian, and Spanish bonds of two, five, or ten-year maturities. $Voice_t^{AN}$ is the vocal sentiment that we derive quantitatively from the vocal emotions of the president (see section during the Q&A session 3.1). An increase in $Voice_t^{AN}$ implies more positive vocal emotions. $Language_t^{AN}$ is the sentiment of individual answers that one can consider as the verbal analog to vocal emotions (for the measurement, see section 3.2). Furthermore, we include an interaction term consisting of the vocal and verbal sentiment of the answers to account for potential non-linear effects arising from the interplay of voice and words.

Even though our main focus is the more situational Q&A session, we do not ignore the role of the introductory statement. Hence, we also include $Voice_t^{IS}$ and $Language_t^{IS}$ to measure the sentiment of vocal and verbal cues of the introductory statement.[10] As additional control variables, we include the forecasts of the ECB/Eurosystem staff as forward-looking indicators ($\sum_{i=6} \beta_i X_{ti}$). The estimation period spans from the end of the acute phase of the sovereign debt crisis in the euro area in July 2013 to the end of Draghi's presidency in October 2019.[11]

[10] As a robustness check (results not shown), we also examined the interaction between vocal and verbal indicators for the introductory statement. The results remain robust, providing no additional insights into what influences asset prices beyond what is presented in this paper.

[11] In principle, we could include the crisis period, but doing so would likely bias the results and raise concerns about a structural break. Additionally, policymakers and observers of monetary policy are generally more interested in the role and interaction of vocal and written reactions from the central bank during non-crisis times.

3.1. Measuring Vocal Sentiment

3.1.1. Design of the Speech Emotion Recognition Model

Quantifying the vocal emotions of the ECB president presents significant challenges. To generate a numerical variable for our event regression estimation, we employ methods from SER, a specialized subfield of machine learning (Pérez-Espinosa et al., 2022). The primary goal of SER is to identify emotions from vocal cues independently of the spoken language. This approach has been recently adopted in economics to assess the impact of vocal sentiment expressed by figures like the Fed chair on asset prices (Gorodnichenko et al., 2023; Alexopoulos et al., 2024). The most relevant study to ours is by Gorodnichenko et al. (2023), where the authors used a Convolutional Neural Network (CNN) to classify the vocal emotions of Federal Reserve chairs during press conferences. Instead, we adopt a Fully Convolutional Neural Network (FCN), as proposed by García-Ordás et al. (2021), due to its superior out-of-sample accuracy[12] and its advantages when handling the dynamic and varied lengths of Q&A session responses.

Unlike traditional CNNs, the FCN model by García-Ordás et al. (2021) is capable of processing audio files of non-fixed length, classifying underlying vocal emotions without the need to pre-process audio into a fixed format.[13] The FCN architecture we use includes three convolutional layers, with the first two using ReLU activation and a dropout layer post the third to prevent overfitting. A pivotal feature of this architecture is the Global Average Pooling layer, which effectively reduces data dimensionality by averaging out filter weights, thus accommodating variable audio lengths without loss of temporal dynamics - a limitation of standard CNNs. This model outputs a feature map for each emotion category, concluding with a Softmax activation layer. The model was constructed using Keras in Python, which

[12]The FCN model demonstrated state-of-the-art out-of-sample accuracy as of August 2021.

[13]For a comprehensive explanation of the FCN model, see the Appendix D and the original article by García-Ordás et al. (2021).

facilitates the implementation of convolutional operations without the constraints imposed by fully connected layers.[14]

The FCN's ability to handle non-fixed length audio data is particularly advantageous. Traditional CNNs, like those used by Gorodnichenko et al. (2023), require audio data to be pre-processed into uniform lengths, often averaging the Mel-frequency cepstral coefficients (MFCC) into a single vector. This process simplifies the input but, more problematically, strips away rich temporal information critical for accurate emotion recognition. In contrast, our FCN approach retains these dynamics, thereby enhancing the detection and classification of nuanced emotional expressions in voice, which can be especially meaningful for the interpretation of monetary policy communications of the kind generated by the ECB.

3.1.2. Model Training and Validation

Following the literature on SER (García-Ordás et al., 2021), and similar to Gorodnichenko et al. (2023), we train and validate our model framework using prepared and labeled emotions using the Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) and the Toronto Emotional Speech Set (TESS). RAVDESS offers 1440 vocal speech emotion audio files generated by 24 actresses and actors (12 female and 12 male) reading two statements with eight different emotions. These emotions are *Neutral, Calm, Happy, Sad, Angry, Fear, Disgust, and Suprised,* and they are available in two different intensities (*normal emotional intensity* and *strong emotional intensity*)[15] (Livingstone and Russo, 2018). TESS contains speech emotion data generated by a young and an old actress, who speak 200 different words with seven different emotions: *Neutral, Happy, Sad, Angry, Fear, Disgust, and (pleasantly) Surprised.* These emotions are considered the "basic emotions" in the neuroscientific literature (Bear et al., 2015, pp. 626 - 628) and are widely used in SER tasks

[14] Appendix A provides a visualization of our FCN model.

[15] For *Neutral* a strong emotional intensity is not available for obvious reasons.

(Pérez-Espinosa et al., 2022). Furthermore, the audio files contain vocal cues recorded in neutral North American English such that trained models are not constrained to English with specific accents (Livingstone and Russo, 2018)[16] which is useful when applying to a proficient English-speaking European.[17]

We follow Gorodnichenko et al. (2023) and remove the emotions *Fear* and *Disgust* due to the low probability that these emotions appear during a central bank press conference.[18] Otherwise, we use all emotions and combine both data sets to a combined emotion set for training and validation of our FCN, that is, we use six emotions for our classification task: *Neutral, Calm, Happy, Sad, Angry, and (pleasantly) Surprised.*[19]

To engineer appropriate features from our emotion set for the FCN, we use the Python package *Librosa* and extract the first 100 Mel-Frequency Cepstrum Coefficients (MFCC) from each audio file.[20] The audio files are not processed or cut in any way. In general, the literature on audio analysis uses different acoustic features, though users decide based on classification accuracy.[21]

For the training and validation of the FCN, we split the emotion set into a training and validation set containing 80% and 20% of the emotions, respectively. To ensure that our

[16] This makes the data more advantageous than vocal data in English with specific accents like in GEMEP.

[17] We do not consider the very few answers that Draghi provides in Italian.

[18] At the beginning of the press conference on 15 April 2015, an activist jumped on the table and disturbed the press conference of the ECB. One could think that this intervention may lead the president to show some fear. Nonetheless, even then, Draghi remained calm leading a journalist to ask: "*[...] And maybe allow me a little add-on question, but due to this intervention of the activist at the beginning of the press conference I observed that you remained very calm. How do you manage this?*"

[19] We also remove, from RAVDESS, sad emotions with high intensity, that is, basically very sad or even crying. After a manual inspection of all Q&A sessions, we can state that Mario Draghi did not cry.

[20] In Appendix B, we explain how we extract MFCCs.

[21] García-Ordás et al. (2021) use and compare Mel Spectrograms with MFCC and conclude that MFCC outperforms the former features regarding classification accuracy. Currently, the literature on speech processing considers that MFCC contains the best characteristics representing the human voice. Nonetheless, a disadvantage of MFCC is their sensitivity to noise (Pérez-Espinosa et al., 2022). Due to this weakness, we cautiously avoid noise and background voices when preparing the actual audio data of Draghi's voice.

model is generalized, that is, the classification is unaffected by the random distribution of emotions into the training and validation set, we use Monte-Carlo Cross Validation (MCCV). We generate seven different training and validation sets, train and validate seven independent FCN models, and combine the models with the highest out-of-sample accuracy into a model ensemble.[22] To ensure that our results are robust to changing stochastic sequences during the training process, we follow the deep learning literature and the literature in economics applying deep learning methods (Gu et al., 2020) and, for every training and validation process, we use a different seed that we generate randomly. To avoid overfitting, we use an "Early Stopping" - Callback.[23]

3.1.3. Using ECB Press Conference Data

Having developed a model framework capable of classifying emotions based on voice, we now apply it to quantify the vocal emotions expressed by the ECB president. We download the audio data of all ECB press conferences—including the Q&A sessions—from the official ECB website (European Central Bank, 2022).[24] We extract only the audio, removing journalists' questions as well as interventions by the Vice-President and the moderator, so that only the ECB president's speech remains.

We examine the introductory statement and the different answers the ECB president provides to journalists. First, we determine the start and end of the introductory statement for each press conference and divide each statement into five-second snippets. This segmen-

[22]Using a model ensemble ensures generalization and reduces the overall classification variance. In Appendix A, table (A.1) shows the out-of-sample accuracy of each FCN model and the average accuracy. The average accuracy is 90.7%.

[23]A common challenge in deep learning is determining the optimal training duration for a model. The 'Early Stopping - Callback' strategy helps address this by halting training before model accuracy and loss on the evaluation sample begin to deteriorate. We use an Intel Core i7 CPU at 2.3 GHz, and the entire training process for the ensemble takes approximately 16 days.

[24]Following a formal request, the ECB provide us written confirmation to use their publicly available video and audio data for research.

tation prevents brief emotional expressions from being averaged out, since using the entire statement would yield only one aggregate emotion and potentially obscure valuable information. Second, we isolate the president's responses during the Q&A sessions. Journalists often pose multiple questions, and while it might seem straightforward to define an answer as the duration between when Draghi begins and ends his response,[25] Draghi often addresses several topics in a single reply, summarizes governing council debates, or responds to multiple questions at once. Treating such long responses as single units again risks averaging out brief vocal signals.

To solve this problem, we exploit an important characteristic in the written transcripts of the ECB press conferences.[26] The ECB staff already identifies the president's individual answers and structures them in separate paragraphs of the transcripts. Therefore, we follow the structure of the ECB press conference transcripts and cut all audio files manually so that the voice in each audio file is identical to the respective paragraph in the transcripts for all press conferences.[27] The output of this process is a novel data set consisting of synchronized voice and language data for future research, which is another contribution of our study.[28] Our voice data consists of 71 introductory statements and 2,336 answers in 71 press conferences between 1 May 2012 and 31 October 2019.[29]

Now, we label the emotions *Happy* and *(pleasantly) Surprised* as **Positive**, *Angry* and *Sad* as

[25]This is the way Gorodnichenko et al. (2023) define the answers of the Fed chair. Nonetheless, this may not be a good choice for the ECB president both because of the possibility of multiple answers and the participation of the ECB Vice-President.

[26]https://www.ecb.europa.eu/press/pressconf/html/index.en.html (last access: 1 August, 2022)

[27]In Appendix C, we provide additional explanation and example of our approach.

[28]We have permission from the ECB confirming that the audio data are not confidential, and the rights of use are publicly available.

[29]Due to poor audio quality, we excluded ten press conferences from our sample: 4 October 2012, 2 May 2013, 2 October 2013, 4 December 2014, 22 October 2015, 2 June 2016, 8 September 2016, 8 December 2016, 8 June 2017, and 6 June 2019.

Negative, and *Neutral* and *Calm* as **Neutral**.[30] We use the voice data of the ECB president, so our model ensemble classifies all statements and answers regarding the underlying vocal sentiment,[31] and we calculate the vocal sentiment for the whole press conference as follows:

$$Voice_t = \frac{Positive_t - Negative_t}{Positive_t + Negative_t} \tag{2}$$

In this formulation, $Positive_t$ and $Negative_t$ represent the number of positive and negative vocal emotions during the press conference at t, respectively. For the introductory statement, they count emotional snippets ($Voice_t^{IS}$); for the Q&A session, they refer to the number of positive and negative answers provided by the ECB president ($Voice_t^{AN}$).[32]

- Figure (1) -

Panel (a) of figure (1) illustrates the development of vocal sentiment during the introductory statements while panel (b) the vocal sentiment during the Q&A sessions. It is plausible that the emotion of the ECB president is strongly affected by monetary policy decision-making, the debates inside the governing council, and the press conference that follows.[33] We observe a correlation of approximately 83% between vocal emotions expressed during the introductory statements and those during the Q&A sessions. On average, the emotional tone during the statements is more positive, with a mean value of 0. This suggests that

[30] Aggregating the emotions in this way has the additional advantage that it improves the classification precision of our FCN model ensemble, due to the similarity of the acoustic features of the emotions aggregated (García-Ordás et al., 2021). Our approach is also consistent with Gorodnichenko et al. (2023) and explicitly considers *calm* emotions.

[31] We use the first six FCN models to classify vocal information, and if they disagree, the seventh FCN model is employed for the final decision.

[32] We use this quantitative measure of vocal emotions for the main analysis. In addition, we provide in Appendix E.2 a robustness check using a qualitative definition of vocal emotions. Our conclusions remain unchanged.

[33] Mayew and Venkatachalam (2012) provide an interesting review of emotions in the psychology literature and emphasize the role of social and interpersonal communication and events that trigger emotions and influence a person's affective state.

changes in emotional state are observed less frequently within a single press conference and more across different press conferences. At the same time, the introductory statement allows for greater emotional control compared to the more situationally driven Q&A sessions.

During the ESDC,[34] vocal sentiment is consistently negative in both the introductory statements and the Q&A sessions. This can be interpreted as a reflection of the intense pressure and psychological stress associated with crisis management.[35] Notably, the press conference on 2 August 2012 - held shortly after Draghi's "Whatever it Takes" speech - shows a temporary spike in sentiment during the introductory statement, marking the most positive moment during the crisis. This moment of vocal optimism stands in contrast to the otherwise somber tone of the period and underscores the significance of Draghi's commitment to preserving the euro.[36]

Following the end of the ESDC, vocal sentiment temporarily improves across both segments, before turning more negative again throughout much of 2014. This period was particularly challenging for the ECB, marked by persistently low inflation, subdued economic growth, growing financial fragility, and concerns about the deanchoring of inflation expectations (Hartmann and Smets, 2018; Rostagno et al., 2021). The launch of the Asset Purchase Programme (APP) in early 2015 is accompanied by a marked increase in positive vocal sentiment—especially pronounced in the introductory statements—possibly reflecting Draghi's success in implementing unconventional monetary policies despite internal dissent within the

[34]We employ the crisis dates as defined by Hartmann and Smets (2018), who identify the acute phase of the ESDC as lasting until June 2013. The authors define a period that starts around the time shortly before the Lehman Brothers' failure (September 2008) until the 'taper tantrum' as the period of financial and sovereign debt crises. The entire euro area sovereign debt crisis is dated May 2010 to June 2013.

[35]As Bernanke (2015) makes clear in his review of the 2008/2009 GFC, enormous psychological stress accompanies crisis management in a financial crisis.

[36]Despite his efforts during the press conference on 2 August 2012, Draghi was heavily criticized for not delivering as much as the markets were expecting, see The Guardian: https://www.theguardian.com/business/blog/2012/aug/02/eurozone-crisis-live-markets-await-ecb-decision (Last access: 4 August 2022). Perhaps in response to this criticism, our model measures a very negative vocal sentiment for the subsequent press conference on 6 September 2012.

governing council (Brunnermeier et al., 2016). However, vocal sentiment begins to decline again in 2018, aligning with rising political and economic uncertainties (Draghi, 2018), and reaches a new low when the ECB resumes its quantitative easing (QE) programme in 2019, just months after attempting to initiate a policy normalization.

3.2. Measuring Verbal Sentiment

Methods from the field of NLP are a well-established part of the methodological toolkit that economists use to analyze central bank communication (Bholat et al., 2015; Benchimol et al., 2022). Additionally, economists employ these methods to detect changes in the language used by central bankers. The literature primarily focuses on the introductory statement and conducts analyses using manually crafted dictionaries (Picault and Renault, 2017; Baranowski et al., 2021; Schmelling and Wagner, 2025). In contrast to previous work, this study introduces two key differences. First, we also examine the more situational Q&A session alongside the introductory statement, as it is of particular interest for vocal analysis. Second, we utilize language modeling to effectively analyze contextual information that dictionary methods can quickly overlook.

To analyze the introductory statement, we utilize the sentiment indicator from Geiger et al. (2025), who derived the indicator by developing a novel and granular artificial intelligence model called MILA. MILA is built on Llama 3.1 70B and is specifically designed and prompted to analyze euro area central bank communication (Deutsche Bundesbank, 2025). The sentiment indicator for the introductory statement is derived through a sentence-by-sentence analysis, considering both in-document context and medium-term inflation expectations as out-of-document context. This sentiment indicator specifically measures the economic optimism or pessimism expressed by the ECB president on behalf of the entire governing council. We use the indicator from Geiger et al. (2025) for $Language_t^{IS}$ because its categories align well with vocal analysis and are derived using similar formulas (Geiger et al., 2025). We use the data from the presidency of Mario Draghi.

To extract the sentiment of verbal cues from the Q&A session, we follow Curti and Kazinnik (2023) and, in line with the findings of Kanelis and Siklos (2025), we use FinBERT, a Transformer-based model developed by Huang et al. (2023). FinBERT excels in discerning sentiment in economic and financial language, adeptly interpreting the tone of complex sentences within their specific context.[37] The analysis happens at the individual sentence level, and we classify each sentence within an answer as either *positive, negative,* or *neutral*. If we count more positive than negative sentences, we classify the answer as on balance positive and vice versa. If an answer only consists of neutral sentences, we classify the answer as neutral. We use the same formula as in the calculation of the vocal sentiment to derive the textual sentiment for each press conference:

$$Language_t^{AN} = \frac{Positive_t - Negative_t}{Positive_t + Negative_t} \tag{3}$$

$Positive_t$ measures the number of positive answers and $Negative_t$ the number of negative answers during the Q&A session of the ECB press conference at t regarding the language. Figure (2) displays the indicator for each press conference. In contrast to the vocal cues during the Q&A sessions, the verbal cues exhibit a higher variation across the press conferences.

- Figure (2) -

Having obtained the time series for vocal and verbal cues for the press conferences, we can now calculate correlations to analyze the relationship of vocal and verbal cues. We provide the correlations in table (1).

[37] Manela and Moreira (2017) provide evidence that automated methods are increasingly superior to lexicographic methods. Huang et al. (2023) show that FinBERT has a higher classification precision than alternative machine learning methods, for example, support vector machines.

- Table (1) -

As previously mentioned, vocal cues during the introductory statement and Q&A session exhibit a high correlation, indicating consistency in Draghi's non-verbal communication behavior during a press conference. Additionally, we observe a positive correlation of nearly 50% between the verbal cues of the introductory statement and the vocal cues of the Q&A session, and a correlation of around 37% for the vocal cues during the Q&A session. The sentiment indicator for the introductory statement reflects the economic optimism or pessimism communicated by the ECB president on behalf of the governing council. A positive correlation suggests that a generally more optimistic view on inflation and economic growth leads to personal optimism or happiness, which is mirrored by a more positive tone of voice. Conversely, we measure a low correlation with the verbal cues during the Q&A session, reflecting the situational and unscripted nature of this part of the conference. This finding aligns with Alexopoulos et al. (2024), who analyzed the non-verbal communication of Fed chairs and found a low correlation between voice and language.

3.3. Control Variables

To control for possible news that Draghi may reveal in parallel with the quarterly ECB / Eurosystem staff projections, we include the difference between the newly published Next Calendar Year (NCY) forecast and the previous NCY forecast for inflation and real GDP growth. To take into account potential asymmetries between positive and negative forecast news, we include variables $\Delta Inflation_t^{NCY,\,Positive}$ and $\Delta RGDP_t^{NCY,\,Positive}$ which equals the difference between the inflation or real GDP forecasts in t and $t-1$ in case the difference is positive and is otherwise zero. Analogously, we include $\Delta Inflation_t^{NCY,\,Negative}$ and $\Delta RGDP_t^{NCY,\,Negative}$ which equals the difference between the inflation or real GDP growth forecast in t and $t-1$ in case the difference is negative and is otherwise zero.

The only known information independent of the ECB press conferences that is often released simultaneously is the weekly US initial jobless claims report. We do not include this information in our regression analysis, as Altavilla et al. (2019) demonstrate that it has only a minimal confounding effect.[38]

4. Empirical Analysis

In our empirical analysis, we concentrate on the period from August 2013 to October 2019. This period encompasses all the press conferences from the formal introduction of FG (Hartmann and Smets, 2018) up to the conclusion of Mario Draghi's tenure as the ECB president. The selection of this specific period is strategic. The formal introduction of FG, as Parle (2022) notes, marked a significant decrease in the volume of unexpected information conveyed during press conferences, thereby diminishing the ECB's informational asymmetry. Moreover, by starting our analysis post-July 2013, we deliberately exclude the ESDC, which was characterized by heightened fiscal instability and the implementation of critical interventions like the Outright Monetary Transactions (OMT) and the European Stability Mechanism (ESM). These factors had a substantial impact on bond yields and could potentially introduce structural breaks in the data. Focusing on the period after these tumultuous events allows for a more controlled analysis, free from the distortions of an acute crisis environment.[39]

---

[38]In a previous version of the paper, we included the monetary policy shocks — Timing, FG, and QE — from Altavilla et al. (2019), as in Möller and Reichmann (2021). However, as Gnan et al. (2025) emphasize, these asset-price-based shocks are not independent of ECB communication but stem from it. As Gnan et al. (2025) note, "researchers can use communication directly instead of having to impose complex sign restrictions to extract news about different aspects of monetary policy indirectly from asset prices." Including both the shocks and our sentiment indicators risks introducing redundancy, as both reflect the same underlying informational content.

[39]When using the full sample, the results and conclusions remain robust for all countries except Italy. This exception likely reflects Italy's unusually volatile period during the ESDC (European Commission, 2013), which skews the mean estimates and may explain the lack of statistical significance. Additionally, we have included a dummy variable for the period until July 2013, and the results remain robust.

4.1. Effects on German Yields

First, we analyze whether the interplay of vocal emotions and language affects the yield of German government bonds, which investors consider to be the benchmark in the euro area (Altavilla et al., 2019).[40] Therefore, we estimate the following regression model, based on equation (1):

$$\begin{aligned} y_t^{DE} = \beta_0 + \beta_1 * Voice_t^{AN} \times Language_t^{AN} + \beta_2 * Voice_t^{AN} + \\ \beta_3 * Language_t^{AN} + \beta_4 * Voice_t^{IS} + \beta_5 * Language_t^{IS} + \\ \sum_{i=6} \beta_i X_{ti} + \epsilon_t \end{aligned} \tag{4}$$

where y_t^{DE} is the median change in the German yield of government bonds with either one-, two-, five-, or ten-year maturities within an interval of ten minutes immediately after the press conference ends. We test the following first hypothesis:

I: *The interplay of vocal emotions and language during the Q&A session influences the yield of sovereign bonds. Positive emotions raise yields as this conveys optimism about current and future expected economic conditions; conversely, negative emotions reduce yields.*

Table (2) shows the estimation results for the yields of German bonds with varying maturities:

- Table (2) around here -

We estimate a significant effect of the verbal cues for the Q&A session ($Language_t^{AN}$) on the yields of bonds with one and two year maturity (see column (1) of table (2)). Indeed, even

[40] When talking about *investors* or *capital markets*, it should be obvious that euro area government bonds are traded by market participants with all kinds of different European and non-European cultural backgrounds and not only by domestic investors of the specific country. This ensures that the effects of emotions are not driven only by country-specific cultural differences in interpreting vocal emotions.

without considering the role of vocal emotions ($Voice_t^{AN}$), a more positive language used by the president to answer the questions leads to an increase of the short-end of the yield curve. As noted above, financial markets perceive the more positive framing of the language and associate it with more optimism about underlying economic conditions. Nonetheless, the sentiment of the introductory statement does not have a similar economic effect which is likely because markets have pretty good expectation on how the statement will be framed. These results align with findings in the literature (Parle, 2022), as discussed in section 2 above.

Interestingly, vocal emotions amplify the effect on German yields. Indeed, omitting a role for vocal emotions cuts the effects of positive cues by roughly one half. Figure (3) provides margin plots for the interaction term such that one can measure the marginal effects of vocal emotions given a fixed level for language ($Positivity_t^{AN}$).[41]

- Figure (3) around here -

More positive vocal cues significantly enhance the impact of verbal cues, resulting in higher yields for all maturities. The voice acts as a vital complement to the verbal cues and framing used by the president, producing effects comparable to monetary tightening.[42] Additionally, for German one-year bonds, the combination of negative vocal and verbal cues results in a decline of yields, aligning with the safe-haven characteristics of these assets.

Finally, increases in the NCY forecasts for inflation lead to rises in the yield curve, consistent with economic theory predictions regarding the direction of the effect. Conversely, the increase in the NCY RGDP forecast reducing short-end yields is somewhat unexpected, as

[41] For illustration, we choose levels of 0.25 and −0.25 for $Positivity_t^{AN}$ in the margin plots since this variable has a mean of −0.07 and a standard deviation of 0.27. Our results remain robust when using different values than 0.25 and −0.25.

[42] This interpretation may be contingent on the downward deviation of the inflation rate from its target. During Draghi's tenure, the primary concern was excessively low inflation. In scenarios where inflation exceeds the target, and monetary policy tightens, positive vocal and verbal cues might have an expansionary effect, as reducing inflation is deemed as due to the successful conduct of monetary policy.

economic theory would typically also anticipate an increase. Moreover, when statistically significant, expectations of inflation and RGDP growth exert a greater influence on yields than other determinants. This is expected, as forecasts, despite their imprecision, provide a numerical indication of the economic outlook.

4.2. Effects on French, Italian, and Spanish Yields

As with the German case, we estimate a significant effect of the interaction term ($Voice_t^{AN} \times Language_t^{AN}$) on the yields of French, Italian, and Spanish government bonds for all maturities. Indeed, Draghi's vocal emotions have a considerably larger impact on Italian yields compared to their French or Spanish counterparts at all maturities. This is likely because international investors are more sensitive to developments in Italy than in other parts of the euro area, reflecting the relative size and fragility of Italy's economy during the sample period considered.[43] Table (3) provides the results.

- Table (3) around here -

Additionally, even with balanced vocal cues, we estimate significant effects from $Language_t^{AN}$ on government bonds in the euro area economies considered. This result clearly indicates that, despite the importance of non-verbal communication, the framing of language remains the primary driver of asset prices at all terms to maturity considered and in all countries examined. More positively framed discussions during the Q&A session lead to increases in yields for all assets. While the results for France and Spain align with those for Germany, the Italian data offer additional insights. For example, German yields are viewed as akin to a relatively risk-free rate, it appears that vocal cues diminish the short-term risk premium in Italy.[44]

---

[43] According to Eurostat, Italy's general government debt-to-GDP ratio has been persistently higher than that of France or Spain. In 2019, Italy's debt ratio was 134.2%, while France's was 97.8% and Spain's was 98.2%. (Eurostat data code is gov_10dd_edpt1 with last access on 07-06-2024.)

[44] Our analysis does not permit conclusions about whether Draghi's vocal cues uniquely impact audiences

Additional asymmetries between Italy and the other countries investigated are obtained by examining the margin plots to measure the effects of vocal emotions given a specific level of positive sentiment in the language used during the Q&A session.[45] Figure (4) shows the marginal effect of vocal emotions on the bonds spreads with two-year maturity:

- Figure (4) around here -

The interplay between vocal emotions and language during the Q&A sessions significantly and asymmetrically influences bond yields, with French and Spanish bond yields moving in one direction and Italian bond yields moving in the opposite direction. For French and Spanish bonds, a combination of positive vocal and verbal cues leads to higher yields, whereas in Italy, a combination of negative vocal and verbal cues results in yield increases. This phenomenon can be attributed to the information channel where unscripted communicative signals are crucial for investors as they assess potential future risks, particularly impacting the risk premium in Italy. These observations align with Mayew and Venkatachalam (2012), who found that investors place greater emphasis on negative over positive vocal emotions in risk assessment. Figure (5) illustrates the marginal effect of vocal emotions on the spreads of government bonds with a five-year maturity:

- Figure (5) around here -

The interplay between vocal emotions and language also impacts the yields of five-year public bonds. For France, yields rise with a combination of positive vocal and verbal cues, while in Italy, they increase when cues are negatively aligned. Clearly, ECB communication signals influencing yields have varying impacts across the euro area's bond markets. Our estimates, however, cannot disentangle the extent to which differences across sovereign bond markets

from different nations. Nonetheless, it's important to acknowledge that these bonds are traded on an international scale, transcending the national identity of the issuing state.

[45] As before, results remain robust when using different values than 0.25 and -0.25 for $Language_t^{AN}$.

are explained by domestic versus international factors. Nevertheless, our results indicate that purely domestic factors, in addition to the euro area-wide determinants considered here, also drive yields.[46] The capital markets prioritize varying information when pricing different public bonds. We reach a similar conclusion for the yields of ten-year French and Spanish bonds, while at the long end of the yield curve, Italian bonds show a similar pattern.[47] Additionally, we find no statistically significant effects from vocal or verbal cues of the introductory statement on yields. This lack of impact is not surprising given the scripted nature of these statements, which allows investors to anticipate them, especially considering the pronounced relevance of FG (Parle, 2022). This result further underscores the importance of examining how unscripted information and signals impact bond yields.

In contrast, the ECB president's unscripted vocal and verbal cues significantly influence public bond yields in the euro area. We observe that both consistently positive and negative cues during Q&A sessions cause bond yields to rise across various countries. This suggests a deeper exploration into how investors react to yield differentials. Initial evidence from Germany suggests that positive communication signals predominantly affect the risk-free yield component. This aligns with the notion that negative signals do not lead to a premium in capital investments viewed as inherently risk-free. Consequently, the next section focuses into euro area yield spreads vis-a-vis German bonds.[48]

---

[46]Since the ECB conducts monetary policy for the euro area, a comprehensive analysis of the role of domestic factors is beyond the scope of this paper.

[47]The margin plots for these ten-year bond estimations are provided in Appendix E.1.

[48]As a robustness check, we replace our numerical variables for measuring vocal sentiment during the introductory statement and the Q&A session with categorical variables based on a Likert scale. We distinguish between *Positive, Moderate Positive, Balanced, Moderate Negative, and Negative.* This scaling aligns more closely with how emotions are perceived in interpersonal relationships. In Appendix E.2, we detail the derivation of these variables and provide the estimations. Our findings remain robust both quantitatively and qualitatively.

4.3. Effects on Spreads

Next, we examine how vocal and verbal cues influence sovereign bond yield differentials to comprehend the effects of unscripted communication on the different components of the yields. Thus, we propose the following hypothesis:

II: *The interplay of vocal emotions and language during the Q&A session influences the spread of French, Italian, and Spanish government bonds. Positive cues reduce yield differentials while negative cues increase them.*

We subtract the yields of French, Italian, and Spanish bonds from the yield of German bonds with identical duration to obtain yield spreads:

$$\overline{y}_t^C = y_t^C - y_t^{DE},\ with\ C \in \{FR, IT, ES\} \tag{5}$$

We use the spread $\overline{y}_t^C$ as the dependent variable and estimate the following regression model based on equation (1):

$$\begin{aligned} \overline{y}_t^C = \beta_0 + \beta_1 * Voice_t^{AN} \times Language_t^{AN} + \beta_2 * Voice_t^{AN} + \beta_3 * Language_t^{AN} + \\ \beta_4 * Voice_t^{IS} + \beta_5 * Language_t^{IS} + \sum_{i=6} \beta_i X_{ti} + \epsilon_t,\ with\ C \in \{FR, IT, ES\} \end{aligned} \tag{6}$$

Figure (6) visualizes the impact of varying levels of vocal cues given a fixed level of verbal cues on the spread by utilizing margin plots:

- Figure (6) around here -

We estimate that consistent negative communication has a statistically and economically significant impact on the spread of Italian bonds. Consequently, negative cues increases the Italian spread and such a result is usually interpreted as a rise in the risk premium. Interestingly, we find no statistically significant effect on the spreads of French and Spanish bonds, leading us to reject the hypothesis for these two countries. Based on the estimation results for yields and spreads, we infer that the increase in French and Spanish yields following consistent positive communication is due to a change in the risk-free interest rate, rather due to a change in the issuer-specific risk premium.

The results reveal a significant asymmetry in the impact of unscripted communication during ECB press conferences. Positive vocal and verbal cues during the Q&A session influence the risk-free interest rate, functioning similarly to a monetary policy impulse. Conversely, negative communication affects the risk premium without impacting the risk-free interest rate. In the appendix, we present the regression results for the spreads of bonds with maturities of five and ten years. These results are qualitatively similar and reinforce our interpretation of the overall yield curve.[49]

In summary, the finding of an asymmetric response to some shocks, whether verbal or economic, is not surprising. However, the differential response of yields at different maturities to ECB announcements and Draghi's verbal emotions suggests that a variable influencing yield spreads may have been omitted.

4.4. Equity Market Response

While the primary focus of this study is on sovereign yields and spreads, we also extend our analysis to two major euro area equity indices — the EURO STOXX 50 (STOXX50) and the EURO STOXX Banks (SX7E) — to assess how vocal and verbal cues from the ECB president impact equity prices in real time. We utilize intra-day data from Altavilla et al. (2019) for stock prices, which is constructed similarly to the sovereign yield data. Figure (7) illustrates the interactive effect of vocal and verbal cues on the stock prices of non-financial and financial European companies.

[49]See Appendix E.1.

- Figure (7) around here -

Notably, negative combinations of vocal and verbal cues correlate with declines in both indices. These findings suggest that negative sentiment conveyed during the Q&A session triggers broader risk-aversion behavior among investors, mirroring the risk-premium increases observed in Italian spreads. Consistent with previous studies (e.g., Mayew and Venkatachalam (2012)), our evidence supports the notion that negative information tends to exert a stronger impact on market participants' risk assessment than positive cues.

5. Conclusions

Our research provides a novel perspective on the impact of non-verbal and unscripted communication on euro area bond yields. We demonstrate that vocal emotions displayed during press conferences significantly influence the pricing of yields and yield spreads of sovereign bonds around the time of ECB governing council meetings. Indeed, and with almost no exception, the interaction of vocal emotions and written text have significantly larger effects at yields across all maturities, and all countries considered, than the other conditioning variables in our regression models.

We highlight the crucial interplay between vocal and verbal cues expressed by the ECB president during the Q&A sessions. Using data from the four largest euro area economies (France, Germany, Italy, and Spain), we establish a statistically significant impact of non-verbal and verbal communication on yields. Specifically, we find that the impact of vocal and verbal communication is the largest for Italy and comparable for France and Spain. Perhaps most importantly, we identify an asymmetric impact of vocal emotions, a previously undocumented finding. In Germany and France, unscripted communication positively affects yields, though this effect is confined to the short end of the yield curve and varies with the type of vocal emotion. Positive communication signals an increase in yields in Germany and

France, while negative verbal and vocal communication leads to increases in the Italian yield curve. Further analysis of bond spreads indicates that negative communication impacts yield spreads. This likely translates into a change in the risk premium vis-a-vis German bonds.

Our results are consistent with the view that investors react to positive signals by anticipating a future monetary policy tightening. Conversely, negative signals affect the risk premium of individual countries, such as Italy, where we observe a relationship between negative communication and increasing spreads. These findings support the hypothesis that euro area bond yields vary idiosyncratically. This provides some support for the often-expressed concern by the ECB over bond market fragmentation in the euro area (e.g., see Ehrmann and Fratzscher (2017)). Therefore, while the ECB conducts a unified monetary policy, sovereign yields can react differently to the central bank's unscripted pronouncements. Nevertheless, our results also demonstrate that euro area monetary policy remains a significant factor in explaining bond spreads of euro area member states relative to German yields.

Our research contributes to the growing body of literature recognizing the influence of vocal cues on financial markets, underscoring that communication extends beyond words alone. That said, much else remains to be explored. Future research should investigate why the same emotions translate into asymmetric effects across sovereign bond markets and how financial markets perceive and process vocal cues during crises, such as the COVID-19 pandemic or the inflation surge of 2021-22. This episode also led to questions about the independence of central banks, and this too may have influenced central bankers' emotions. Understanding how central bankers' emotions influence asset prices amid increasing economic and geopolitical uncertainty represents another promising avenue for further investigation. More challenging will be how to properly control for differences in individual personalities and gender who lead central banks. Given the rate of technical progress in deciphering and measuring the content of vocal communications it is also likely that greater precision in estimating their impact will be possible and new insights obtained. We leave these extensions for future research.

References

Alexopoulos, M., Han, X., Kryvtsov, O., Zhang, X., 2024. *More than Words: Fed Chairs' Communication During Congressional Testimonies*. Journal of Monetary Economics 142.

Altavilla, C., Brugnolini, L., Gürkaynak, R.S., Motto, R., Ragusa, G., 2019. *Measuring Euro Area Monetary Policy*. Journal of Monetary Economics 108, 162–179.

Apel, M., Blix Grimaldi, M., 2014. *How Informative Are Central Bank Minutes?* Review of Economics 65, 53–76.

Baranowski, P., Bennani, H., Doryn, W., 2021. *Do the ECB's Introductory Statements Help Predict Monetary Policy? Evidence from a Tone Analysis*. European Journal of Political Economy 66.

Barry, M., Bruns, B., Kandemir, S., Klose, J., Smirnov, V., Tillmann, P., 2025. *The Emotions of Monetary Policy*. Working Paper .

Bear, M.F., Connors, B., Paradiso, M., 2015. *Neuroscience: Exploring the Brain.* Fourth, international ed., Jones & Bartlett Learning.

Benchimol, J., Kazinnik, S., Saadon, Y., 2022. *Text Mining Methodologies with R: An Application to Central bank Texts*. Machine Learning with Applications 8.

Bernanke, B., 2015. *The Courage to Act: A Memoir of a Crisis and Its Aftermath.* 1 ed., WW Norton.

Bholat, D., Hansen, S., Santos, P., Schonhardt-Bailey, C., 2015. *Text Mining for Central Banks*. Centre for Central Banking Studies Handbook No. 33.

Blinder, A.S., Ehrmann, M., De Haan, J., Jansen, D.J., 2024. *Central Bank Communication with the General Public: Promise or False Hope?* Journal of Economic Literature Forthcoming.

Blinder, A.S., Ehrmann, M., Fratzscher, M., De Haan, J., Jansen, D.J., 2008. *Central Bank Communication and Monetary Policy: A Survey of Theory and Evidence.* Journal of Economic Literature 46, 910–45.

Bohl, M.T., Kanelis, D., Siklos, P.L., 2023. *Central Bank Mandates: How Differences can Influence the Content and Tone of Central Bank Communication.* Journal of International Money and Finance 130, 102752.

Brand, C., Buncic, D., Turunen, J., 2010. *The Impact of ECB Monetary Policy Decisions and Communication on the Yield Curve*. Journal of the European Economic Association 8(6), 1266–1298.

Brunnermeier, M.K., James, H., Landau, J.P., 2016. *The Euro and the Battle of Ideas.* Princeton University Press.

Cieslak, A., Schrimpf, A., 2019. *Non-Monetary News in Central Bank Communication*. Journal of International Economics 118, 293–315.

Cochrane, J.H., Piazzesi, M., 2002. *The FED and Interest Rates - A High-Frequency Identification*. American Economic Review 92(2), 90–95.

Curti, F., Kazinnik, S., 2023. *Let's Face It: Quantifying the Impact of Nonverbal Communication in FOMC Press Conferences*. Journal of Monetary Economics 139, 110–126.

Deutsche Bundesbank, 2025. *Monetary Policy Communication According to Artificial Intelligence*. Monthly Report - March 2025 .

Draghi, M., 2018. *The Outlook for the Euro Area Economy*. Speech by Mario Draghi, President of the ECB, Frankfurt European Banking Congress, Frankfurt am Main, 16 November 2018.

Ehrmann, M., Fratzscher, M., 2017. *Euro area government bonds – Fragmentation and contagion during the sovereign debt crisis*. Journal of International Money and Finance 70, 26–44.

Ehrmann, M., Talmi, J., 2020. *Starting from a Blank Page? Semantic Similarity in Central Bank Communication and Market Volatility*. Journal of Monetary Economics 111, 48–62.

Ehrmann, M., Wabitsch, A., 2022. *Central Bank Communication with Non-Experts - A Road to Nowhere?* Journal of Monetary Economics 127, 69–85.

European Central Bank, 2022. *Webcasts of ECB Press Conferences*. Retrived from: https://www.ecb.europa.eu/press/tvservices/webcast/html/index.en.html. Last Access: 30 May 2022.

European Commission, 2013. *European Economy: Macroeconomic Imbalances Italy 2013*. Occasional Paper Series April 2013. Last Access: 30 May 2022.

Ferrara, F.M., Angino, S., 2022. *Does Clarity Make Central Banks More Engaging? Lessons From ECB Communications*. European Journal of Political Economy 74. 102146.

Filippi, P., Ocklenburg, S., Bowling, D., Heege, L., Güntürkün, O., Newen, A., De Boer, B., 2017. *More than words (and faces): evidence for a Stroop effect of prosody in emotion word processing*. Cognition and Emotion 31.

García-Ordás, M.T., Alaiz-Moretón, H., Benítez-Andrades, J.A., García-Rodríguez, I., García-Olalla, O., Benavides, C., 2021. *Sentiment Analysis in Non-Fixed Length Audios Using a Fully Convolutional Neural Network*. Biomedical Signal Processing and Control 69.

Geiger, F., Kanelis, D., Lieberknecht, P., Sola, D., 2025. *Monetary-Intelligent Language Agent*. Deutsche Bundesbank Technical Paper Series .

Gnan, P., Schleritzko, M., Schmeling, M., Wagner, C., 2025. *Deciphering Monetary Policy Shocks*. Working Paper SSRN.

Gorodnichenko, Y., Pham, T., Talavera, O., 2023. *The Voice of Monetary Policy*. American Economic

Review Vol. 113, No. 2.

Gu, S., Kelly, B., Xiu, D., 2020. *Empirical Asset Pricing via Machine Learning*. Review of Financial Studies , 2223–2273.

Gürkaynak, R.S., Sack, B., Swanson, E.T., 2005. *Do Actions Speak Louder Than Words? The Response of Asset Prices to Monetary Policy Actions and Statements*. International Journal of Central Banking , 55–93.

Hartmann, P., Smets, F., 2018. *The European Central Bank's Monetary Policy during Its First 20 Years*. Brookings Papers on Economic Activity, Fall, 1-146.

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

Huang, A.H., Wang, H., Yang, Y., 2023. *FinBERT - A Large Language Model Approach to Extracting Information from Financial Text*. Contemporary Accounting Research 40(2).

Hubert, P., Labondance, F., 2021. *The Signaling Effects of Central Bank Tone*. European Economic Review 133.

Kanelis, D., Siklos, P.L., 2025. *The ECB Press Conference Statement: Deriving a New Sentiment Indicator for the Euro Area.* International Journal of Finance & Economics 30, 652–664.

Kuttner, K.N., 2001. *Monetary Policy Surprises and Interest Rates: Evidence from the Fed Funds Future Market*. Journal of Monetary Economics 47, 523–544.

Livingstone, S., Russo, F., 2018. *The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English.* PLoS One 5.

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

Manela, A., Moreira, A., 2017. *News Implied Volatility and Disaster Concerns*. Journal of Financial Economics 123(1), 137–162.

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

Mehrabian, A., 1971. *Silent Messages: Implicit Communication of Emotions and Attitudes.* Belmont, CA: Wadsworth.

Möller, R., Reichmann, D., 2021. *ECB language and stock returns - A textual analysis of ECB press conferences*. The Quarterly Review of Economics and Finance 80, 590–604.

Parle, C., 2022. *The Financial Market Impact of ECB Monetary Policy Press Conferences - A Text Based Approach*. European Journal of Political Economy 74. 102230.

Picault, M., Renault, T., 2017. *Words are not all Created Equal: A New Measure of ECB Communication*. Journal of International Money and Finance 79, 136–156.

Pérez-Espinosa, H., Zatarain-Cabada, R., Barrón-Estrada, M.L., 2022. *Emotion Recognition: From Speech and Facial Expressions.* chapter 15. Published In: Torres-García, Alejandro A. and Reyes-García, Carlos A. and Villaseñor-Pineda, Luis and Mendoza-Montoya, Omar (Ebs.): Biosignal Processing and Classification Using Computational Learning and Intelligence. Elsevier.

Ramey, V.A., 2016. *Macroeconomic Shocks and Their Propagration.* chapter 2. Published In: Taylor, John B. and Uhlig, Harald (Ebs.): Handbook of Macroeconomics. Volume 2. Elsevier.

Rigobon, R., Sack, B., 2004. *The Impact of Monetary Policy on Asset Prices*. Journal of Monetary Economics 51, 1553–1575.

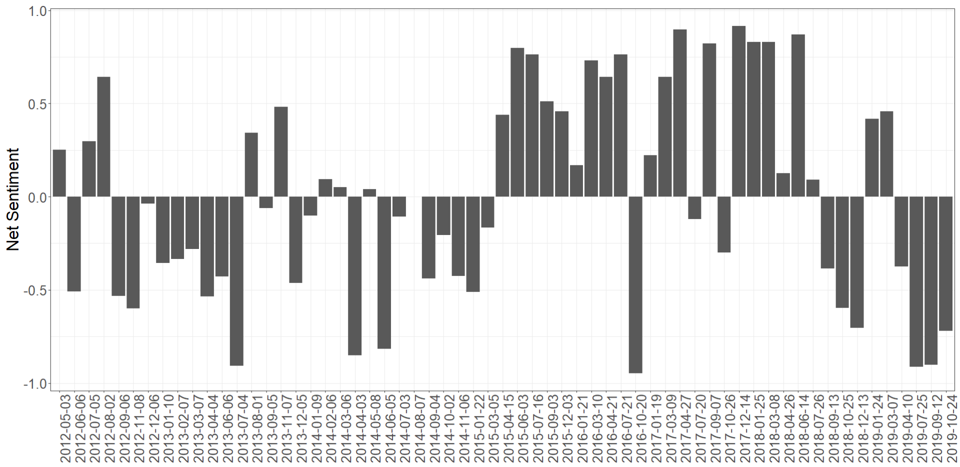
Rostagno, M., Altavilla, C., Carboni, G., Lemke, W., Motto, R., Guilhem, A.S., Yiangou, J., 2021. *Monetary Policy in Times of Crisis: A Tale of Two Decades of the European Central Bank.* 1 ed., Oxford University Press.

Schmelling, M., Wagner, C., 2025. *Does Central Bank Tone Move Asset Prices?* Journal of Financial and Quantitative Analysis 60(1), 36–67.

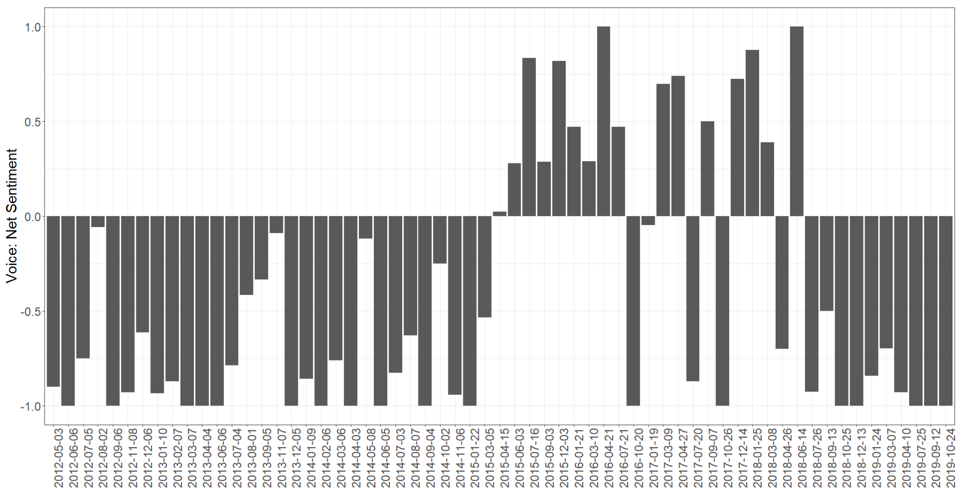
Shapiro, A.H., Wilson, D., 2021. *Taking the Fed at its Word: A New Approach to Estimating Central Bank Objectives using Text Analysis*. Review of Economic Studies .

Swanson, E.T., 2021. *Measuring the Effects of Federal Reserve Forward Guidance and Asset Purchases on Financial Markets*. Journal of Monetary Economics 118, 32–53.

Figure 1: Sentiment of vocal cues during the ECB Press Conferences



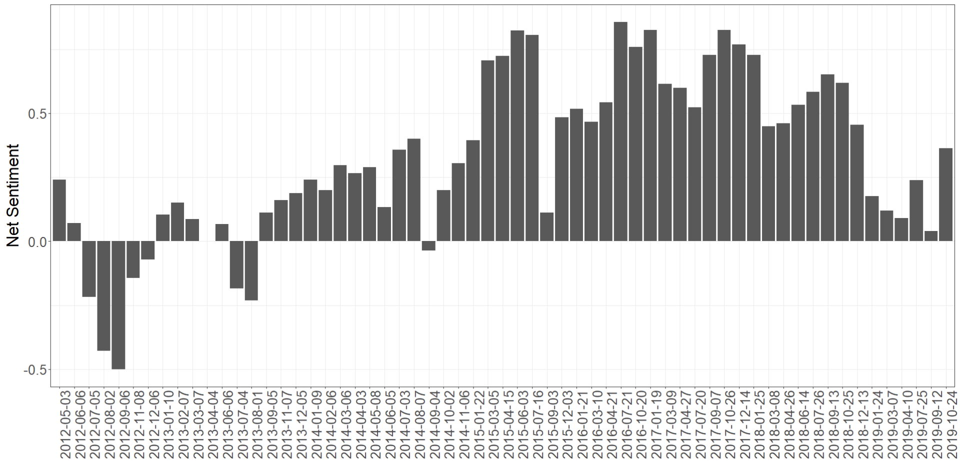
(a) Sentiment of vocal cues during introductory statements



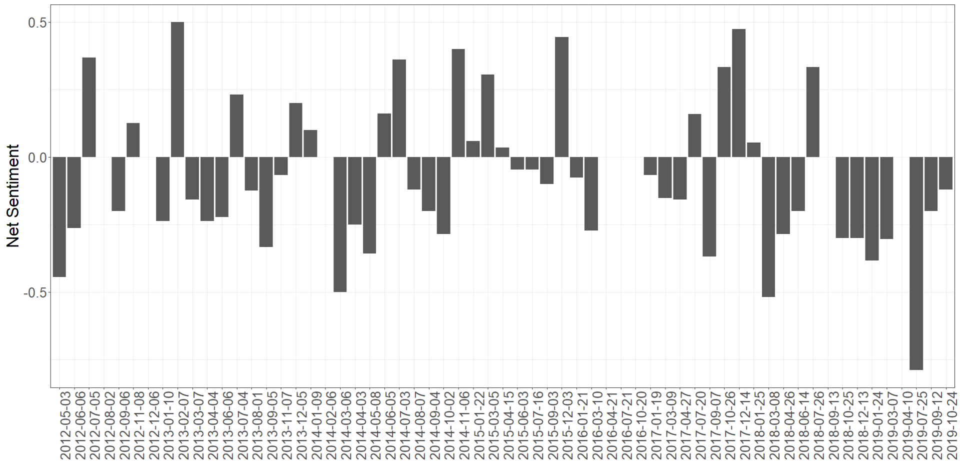
(b) Sentiment of vocal cues during Q&A sessions

Note: Panel (a) illustrates the sentiment of the vocal emotions measurable in Mario Draghi's voice during the introductory statement press conferences. Panel (b) illustrates the sentiment of the vocal emotions measurable in Mario Draghi's answers during the Q&A session of the press conferences. We measure vocal sentiment using our SER model (see section 3.1) and calculate the net value using equation (2).

Figure 2: Sentiment of verbal cues during the ECB Press Conferences



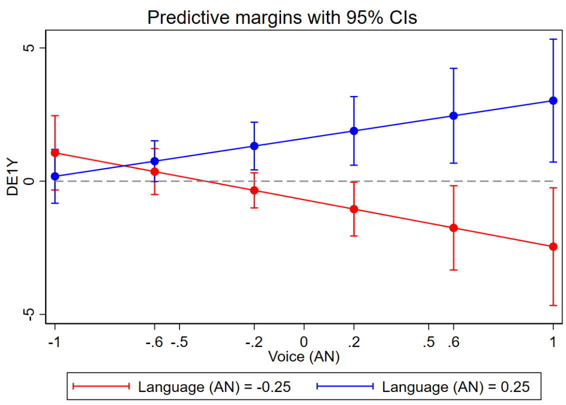
(a) Sentiment of verbal cues during introductory statement



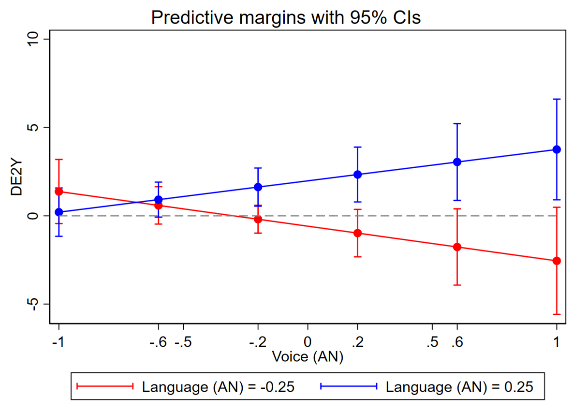
(b) Sentiment of verbal cues during Q&A sessions

Note: Panel (a) illustrates the verbal sentiment of the introductory statement using the data from Geiger et al. (2025). Panel (b) illustrates the verbal sentiment of Mario Draghi's answers during the Q&A session of the ECB press conferences. We measure verbal sentiment during the Q&A using FinBERT (see section 3.2) and calculate the value using equation (3).

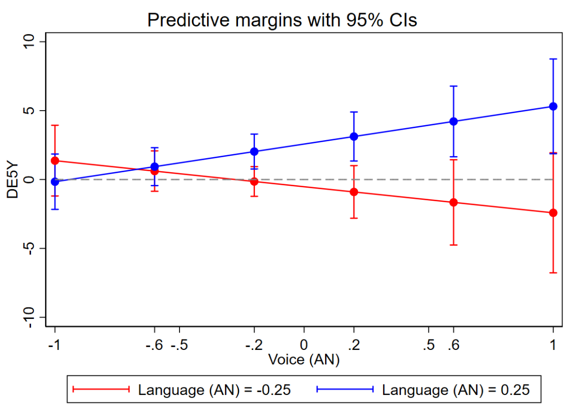
Figure 3: Marginal Effect of Vocal Emotions Given Language on German Yields



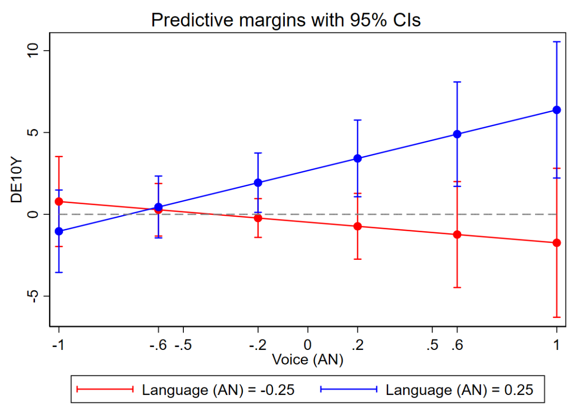
(a) One-Year German Yield



(b) Two-Year German Yield



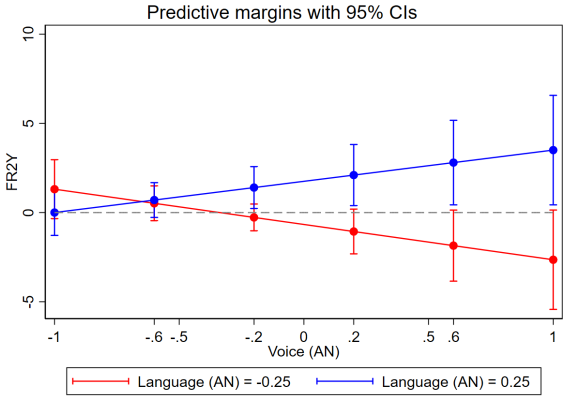
(c) Five-Year German Yield



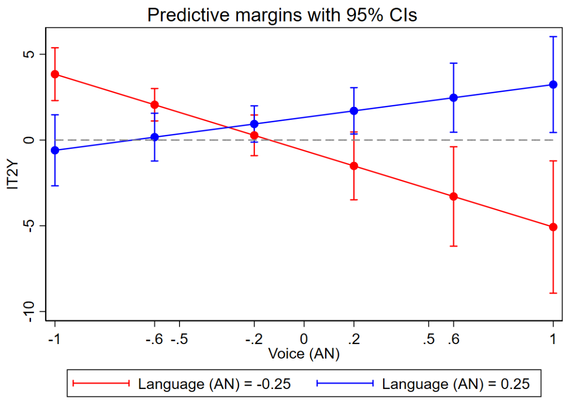
(d) Ten-Year German Yield

Note: These plots visualize the marginal effect of a change in $Voice_t$ given a specific level of $Language_t^{AN}$ on the yields of German government bonds. We report the change in yields in basis points. We illustrate the marginal effect of $Voice_t^{AN}$ for $Language_t^{AN} \in \{-0.25,\ 0.25\}$. We report the estimations in table (2).

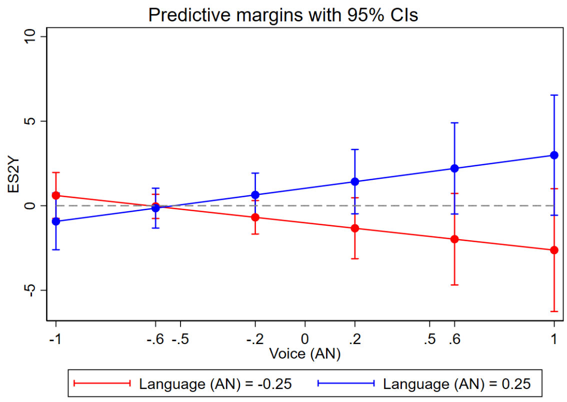
Figure 4: Marginal Effect of Vocal Emotions Given Language on Two-Year Yields



(a) Two-Year French Yield



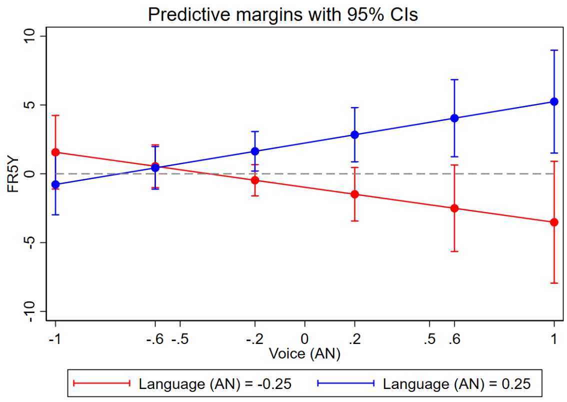
(b) Two-Year Italian Yield



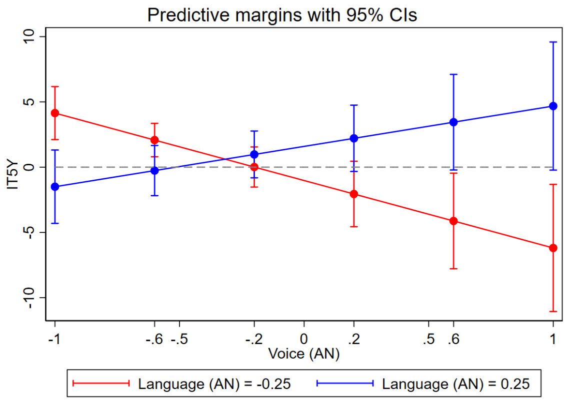
(c) Two-Year Spanish Yield

Note: These plots visualize the marginal effect of a change in $Voice_t^{AN}$ given a specific level of $Language_t^{AN}$ on the yield of two-year government bonds from a) France, b) Italy, and c) Spain. We report the change in yields in basis points. We illustrate the marginal effect of $Voice_t^{AN}$ for $Language_t^{AN} \in \{-0.25,\ 0.25\}$. We report the estimations in table (3).

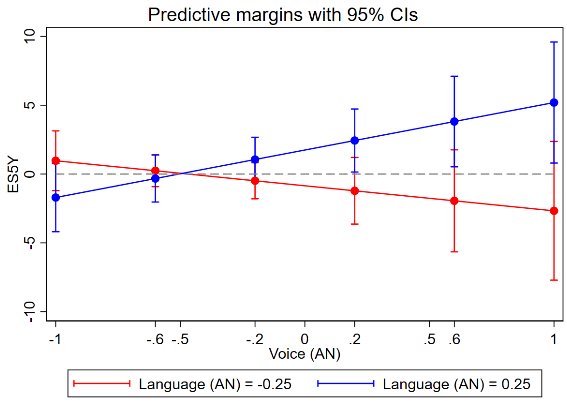
Figure 5: Marginal Effect of Vocal Emotions Given Language on Five-Year Yields



(a) Five-Year French Yield



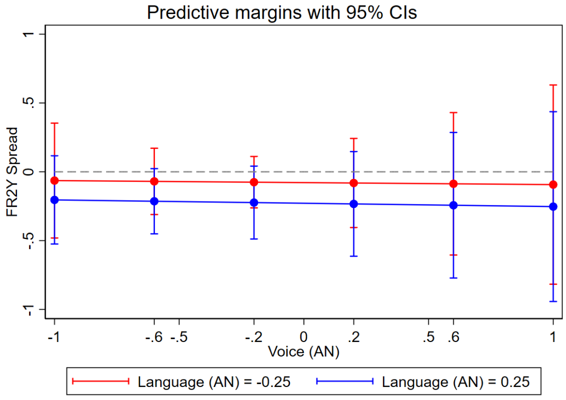
(b) Five-Year Italian Yield



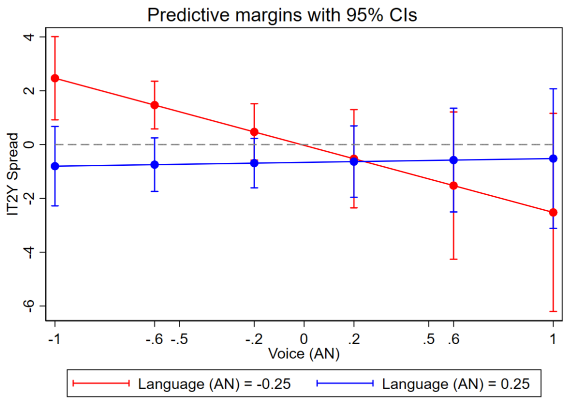
(c) Five-Year Spanish Yield

Note: These plots visualize the marginal effect of a change in $Voice_t^{AN}$ given a specific level of $Language_t^{AN}$ on the yields of five-year government bonds from a) France, b) Italy, and c) Spain. We report the change in yields in basis points. We illustrate the marginal effect of $Voice_t^{AN}$ for $Language_t^{AN} \in \{-0.25,\ 0.25\}$. We report the estimations in table (3).

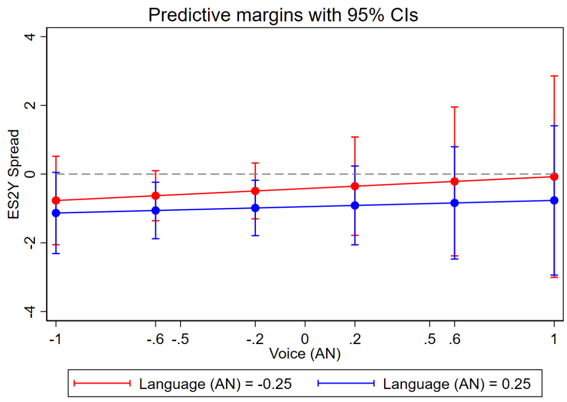
Figure 6: Marginal Effect of Vocal Emotions Given Language on Two-Year Spreads



(a) Two-Year French Spread



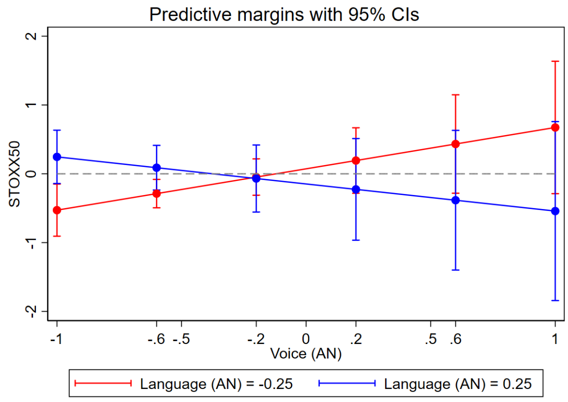
(b) Two-Year Italian Spread



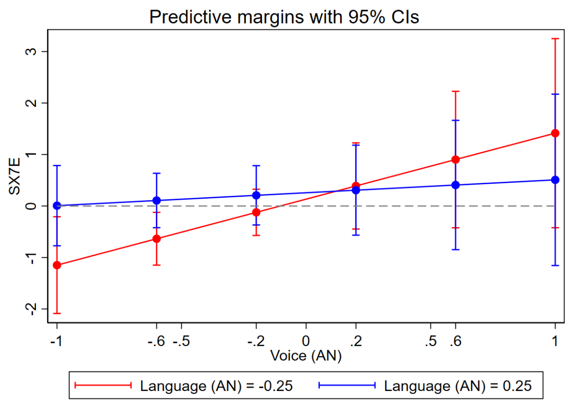
(c) Two-Year Spanish Spread

Note: These plots visualize the marginal effect of a change in $Voice_t^{AN}$ given a specific level of $Language_t^{AN}$ on the spread of two-year government bonds from a) France, b) Italy, and c) Spain. We report the change in yields in basis points. We illustrate the marginal effect of $Voice_t^{AN}$ for $Language_t^{AN} \in \{-0.25,\ 0.25\}$.

Figure 7: Marginal Effect of Vocal Emotions Given Language on Stock Prices



(a) EURO STOXX 50



(b) EURO STOXX Banks

Note: These plots visualize the marginal effect of a change in $Voice_t^{AN}$ given a specific level of $Language_t^{AN}$ on the EURO STOXX 50 and EURO STOXX Banks. We report the change in basis points. We illustrate the marginal effect of $Voice_t^{AN}$ for $Language_t^{AN} \in \{-0.25,\ 0.25\}$.

Table 1: **Correlations of Vocal and Verbal Cues**

	$Voice_t^{IS}$	$Voice_t^{AN}$	$Positivity_t^{IS}$	$Positivity_t^{AN}$
$Voice_t^{IS}$	1.000			
$Voice_t^{AN}$	0.834	1.000		
$Positivity_t^{IS}$	0.367	0.478	1.000	
$Positivity_t^{AN}$	0.003	0.023	0.112	1.000

Notes: This table presents the correlation coefficients for the sentiment of vocal cues during the introductory statements ($Voice^{IS}t$) and Q&A sessions ($Voice^{AN}t$), as well as the sentiment of verbal cues during the introductory statements ($Positivity^{IS}t$) and Q&A sessions ($Positivity^{AN}t$).

Table 2: **German Bond Yields**

	(1)	(2)	(3)	(4)
Dependent Variable	DE1Y	DE2Y	DE5Y	DE10Y
$Voice_t^{AN} \times Language_t^{AN}$	6.36***	7.47***	9.26***	9.94***
	(1.40)	(1.84)	(2.27)	(2.95)
$Voice_t^{AN}$	-0.17	-0.09	0.42	1.22
	(0.68)	(0.91)	(1.30)	(1.36)
$Language_t^{AN}$	4.60***	5.14***	6.20	6.31
	(1.40)	(1.58)	(1.70)	(2.29)
$Voice_t^{IS}$	0.27	0.49	0.70	0.42
	(0.83)	(1.12)	(1.61)	(1.806)
$Language_t^{IS}$	-1.12	-1.59	-2.83	-3.16
	(0.96)	(1.55)	(2.11)	(2.21)
$\Delta Inflation_t^{NCY,Positive}$	8.77***	11.60**	12.04**	5.36
	(3.13)	(4.76)	(5.62)	(5.19)
$\Delta Inflation_t^{NCY,Negative}$	7.96	8.16	8.91	8.64
	(5.31)	(6.51)	(8.90)	(8.62)
$\Delta RGDP_t^{NCY,Positive}$	-7.61**	-10.08**	-13.79**	-9.30*
	(3.27)	(4.25)	(5.06)	(5.29)
$\Delta RGDP_t^{NCY,Negative}$	1.03	-2.15	-8.87	-5.90
	(3.52)	(4.22)	(6.18)	(5.96)
Constant	0.57	1.06	2.20	2.46
	(3.52)	(0.22)	(1.39)	(1.46)
R^2	0.48	0.43	0.41	0.34
Obs	48	48	48	48

Note: Results are based on equation (1) using the change in yield of one, two, five, and ten-year government bonds of Germany for August 2013 until October 2019 as the dependent variable. The dependent variable is calculated as the difference in the median price in a narrow time window before the ECB press conference and a narrow time window afterward and is reported in basis points. We report the change in yields in basis points. $Voice_t^{IS}$ and $Voice_t^{AN}$ are Draghi's vocal sentiment during the introductory statement and Q&A session, respectively (see 3.1). Analogously, $Language_t^{IS}$ and $Language_t^{AN}$ measure the sentiment of Draghi's verbal cues. Furthermore, $\Delta Inflation^{NCY,Positive}$ ($\Delta RGDP^{NCY,Positive}$) controls for the change in the NCY forecast for inflation (real GDP) from the ECB/Eurosystem staff projections. Analogously, $\Delta Inflation^{NCY,Negative}$ ($\Delta RGDP^{NCY,Negative}$) controls for the effects of negative changes. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: **French, Italian, and Spanish Bond Yields**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent Variable	FR2Y	IT2Y	ES2Y	FR5Y	IT5Y	ES5Y	FR10Y	IT10Y	ES10Y
$Voice_t^{AN} \times Language_t^{AN}$	7.46***	12.75***	7.15**	11.10***	16.51***	10.54***	13.71***	20.60***	12.99***
	(1.98)	(2.61)	(2.64)	(2.48)	(3.91)	(3.28)	(3.16)	(4.39)	(3.90)
$Voice_t^{AN}$	-0.11	-1.27	0.17	0.23	-1.04	0.82	0.92	0.82	0.82
	(0.84)	(0.96)	(0.95)	(1.35)	(1.30)	(1.37)	(1.57)	(1.68)	(1.71)
$Language_t^{AN}$	4.84**	3.87**	4.08**	6.43***	5.23*	5.20**	7.81***	7.68**	6.51**
	(1.78)	(1.38)	(1.92)	(1.93)	(2.66)	(2.39)	(2.43)	(3.28)	(3.10)
$Voice_t^{IS}$	0.52	0.85	0.78	0.83	1.45	-0.45	0.73	0.95	0.53
	(1.02)	(1.41)	(1.08)	(1.70)	(1.67)	(1.68)	(2.09)	(2.18)	(2.15)
$Language_t^{IS}$	-1.20	1.50	-1.23	-1.39	1.61	-0.70	-1.87	0.48	-0.31
	(1.32)	(1.46)	(1.37)	(2.27)	(2.14)	(1.72)	(2.41)	(2.53)	(2.57)
$\Delta Inflation_t^{NCY,Positive}$	10.95**	12.23**	7.46*	12.41**	11.40**	9.97**	9.33	11.29*	8.02
	(4.03)	(4.85)	(4.01)	(5.41)	(5.36)	(4.60)	(6.02)	(6.18)	(5.93)
$\Delta Inflation_t^{NCY,Negative}$	6.87	14.62	-1.46	9.55	3.81	1.62	7.77	7.06	3.39
	(6.19)	(11.50)	(6.42)	(8.92)	(10.42)	(9.32)	(9.10)	(10.96)	(9.03)
$\Delta RGDP_t^{NCY,Positive}$	-10.11**	-15.60***	-9.14**	-14.10**	-16.41***	-12.43**	-11.85**	-17.03***	-13.30**
	(3.88)	(4.46)	(4.06)	(5.01)	(5.31)	(4.77)	(5.63)	(5.95)	(5.66)
$\Delta RGDP_t^{NCY,Negative}$	1.15	1.53	-3.66	-5.64	-2.00	2.17	-1.99	-3.18	-3.23
	(4.21)	(6.44)	(5.60)	(5.98)	(8.42)	(6.96)	(5.98)	(8.90)	(7.43)
Constant	0.72	-0.73	0.70	1.07	-0.36	0.77	1.58	0.99	1.04
	(0.88)	(1.20)	(1.15)	(1.47)	(1.62)	(1.48)	(1.60)	(1.86)	(1.85)
R^2	0.423	0.471	0.305	0.392	0.429	0.302	0.358	0.471	0.429
Obs	48	48	48	48	47	48	48	48	48

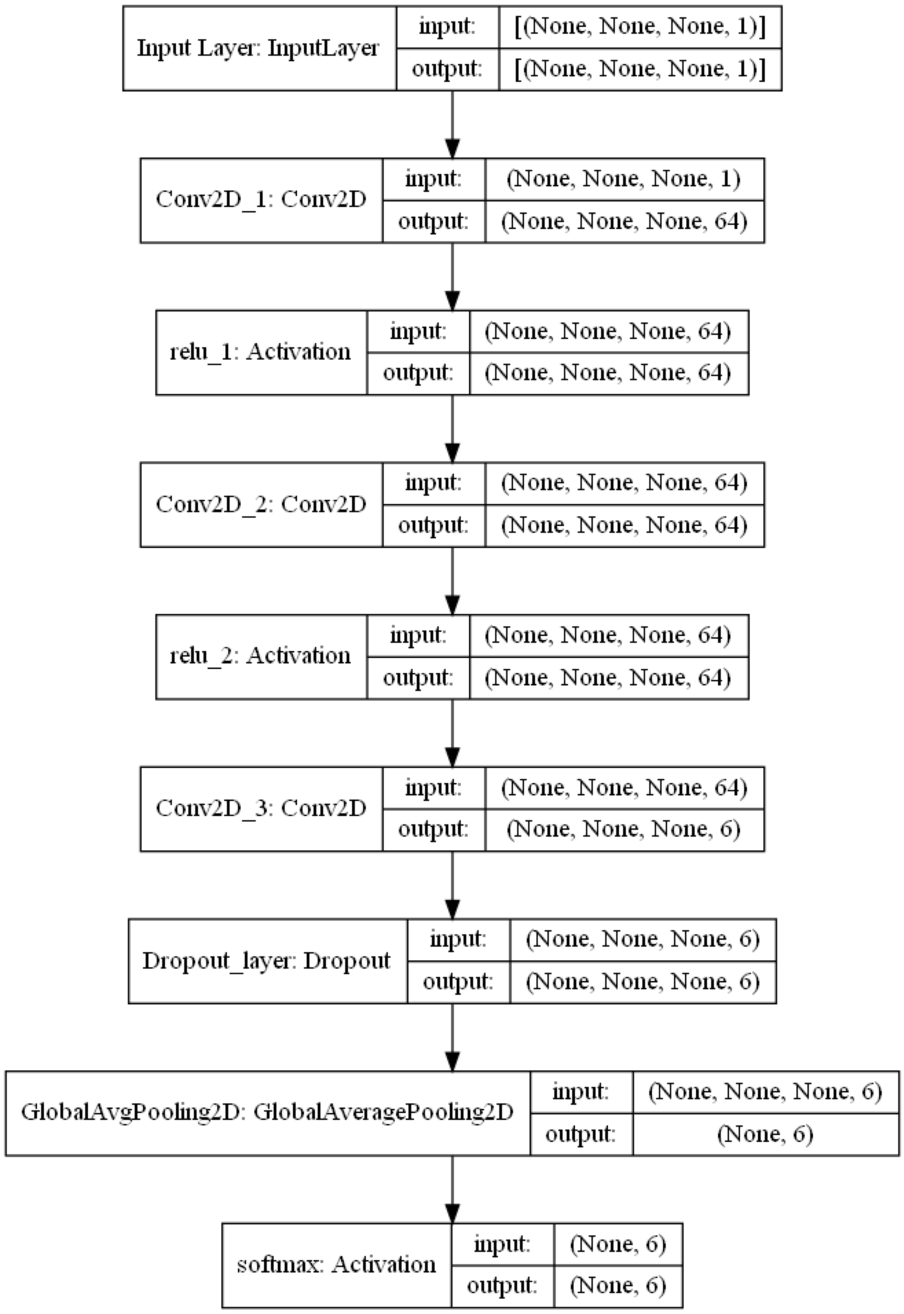
Note: Results are based on equation (1) using the change in the yield of two, five, and ten-year government bonds of France, Italy, and Spain for August 2013 until October 2019 as the dependent variable. The dependent variable is calculated as the difference in the median price in a narrow time window before the ECB press conference and a narrow time window afterward. We report the change in yields in basis points. $Voice_t^{IS}$ and $Voice_t^{AN}$ are Draghi's vocal sentiment during the introductory statement and Q&A session, respectively (see 3.1). Analogously, $Language_t^{IS}$ and $Language_t^{AN}$ measure the sentiment of Draghi's verbal cues. Furthermore, $\Delta Inflation^{NCY,Positive}$ ($\Delta RGDP^{NCY,Positive}$) controls for the change in the NCY forecast for inflation (real GDP) from the ECB/Eurosystem staff projections. Analogously, $\Delta Inflation^{NCY,Negative}$ ($\Delta RGDP^{NCY,Negative}$) controls for the effects of negative changes. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Online Appendix to "Emotion in Euro Area Monetary Policy Communication and Asset Prices: The Draghi Era"

Dimitrios Kanelis Pierre L. Siklos

Appendix A. Additional Figures and Tables

Figure A1: FCN Model Representation



Note: This figure provides a representation of the FCN model we are using for the SER.

Table A.1: **FCN Out-of-Sample Classification Precision**

1. FCN	**2. FCN**	**3. FCN**	**4. FCN**	**5. FCN**	**6. FCN**	**7. FCN**	**Average**
92.7%	90.9%	91.6%	90.5%	88.7%	90.5%	89.9%	90.7%

Note: This table displays the out-of-sample accuracy of each FCN model that achieved the highest precision in its respective training session. Identical hyperparameters were used across sessions, but the composition of training and validation sets, as well as the randomization seed, varied to enhance generalization.

Table A.2: **Vocal and Verbal Sentiment Data (Part I)**

Date	$Voice_t^{IS}$	$Language_t^{IS}$	$Voice_t^{AN}$	$Language_t^{AN}$
03-05-2012	0.25	0.24	-0.90	-0.44
06-06-2012	-0.51	0.07	-1.00	-0.26
05-07-2012	0.30	-0.22	-0.75	0.37
02-08-2012	0.64	-0.43	-0.06	0
06-09-2012	-0.53	-0.50	-1.00	-0.20
08-11-2012	-0.60	-0.14	-0.93	0.125
06-12-2012	-0.04	-0.07	-0.61	0.00
10-01-2013	-0.36	0.10	-0.94	-0.24
07-02-2013	-0.33	0.15	-0.87	0.5
07-03-2013	-0.28	0.09	-1.00	-0.16
04-04-2013	-0.53	0.00	-1.00	-0.24
06-06-2013	-0.43	0.07	-1.00	-0.22
04-07-2013	-0.91	-0.19	-0.79	0.23
01-08-2013	0.34	-0.23	-0.42	-0.13
05-09-2013	-0.06	0.11	-0.33	-0.33
07-11-2013	0.48	0.16	-0.09	-0.07
05-12-2013	-0.46	0.19	-1.00	0.20
09-01-2014	-0.10	0.24	-0.86	0.10
06-02-2014	0.10	0.20	-1.00	0.00
06-03-2014	0.05	0.30	-0.76	-0.50
03-04-2014	-0.85	0.27	-1.00	-0.25
08-05-2014	0.04	0.29	-0.12	-0.36
05-06-2014	-0.82	0.13	-1.00	0.16
03-07-2014	-0.11	0.36	-0.83	0.36
07-08-2014	0.00	0.40	-0.63	-0.12
04-09-2014	-0.44	-0.04	-1.00	-0.20
02-10-2014	-0.21	0.2	-0.25	-0.29
06-11-2014	-0.43	-0.04	-1.00	-0.20

Note: This table presents the time series for vocal cues during the introductory statement ($Voice_t^{IS}$) and Q&A session ($Voice_t^{AN}$), as well as verbal cues during the introductory statement ($Language_t^{IS}$) and Q&A session ($Language_t^{AN}$).

Table A.3: **Vocal and Verbal Sentiment Data (Part II)**

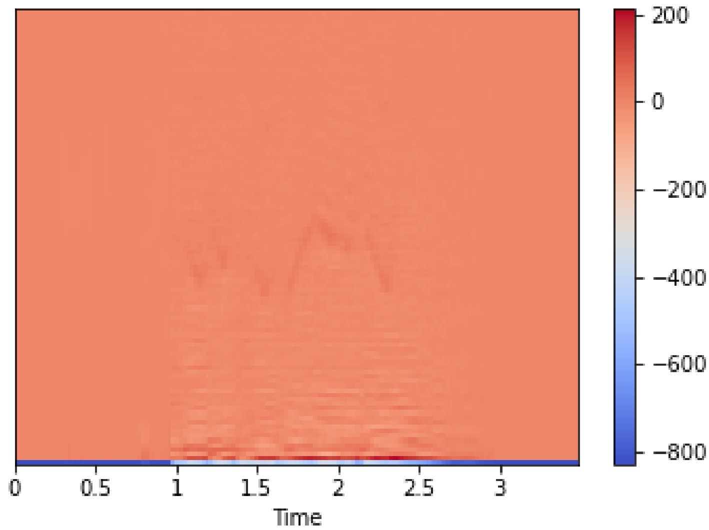
Date	$Voice_t^{IS}$	$Language_t^{IS}$	$Voice_t^{AN}$	$Language_t^{AN}$
22-01-2015	-0.51	0.39	-1.00	0.06
05-03-2015	-0.17	0.71	-0.53	0.30
15-04-2015	0.44	0.72	0.02	0.03
03-06-2015	0.80	0.82	0.28	-0.05
16-07-2015	0.76	0.81	0.83	-0.05
03-09-2015	0.51	0.11	0.29	-0.10
03-12-2015	0.46	0.48	0.82	0.44
21-01-2016	0.17	0.52	0.47	-0.08
10-03-2016	0.73	0.47	0.29	-0.27
21-04-2016	0.64	0.54	1.00	0.00
21-07-2016	0.76	0.86	0.47	0.00
20-10-2016	-0.95	0.76	-1.00	0.00
19-01-2017	0.22	0.83	-0.05	-0.07
09-03-2017	0.64	0.62	0.70	-0.15
27-04-2017	0.90	0.60	0.74	-0.16
20-07-2017	-0.12	0.52	-0.87	0.16
07-09-2017	0.82	0.73	0.50	-0.37
26-10-2017	-0.30	0.83	-1.00	0.33
14-12-2017	0.92	0.77	0.72	0.47
25-01-2018	0.83	0.73	0.88	0.05
08-03-2018	0.83	0.45	0.39	-0.52
26-04-2018	0.13	0.46	-0.70	-0.29
14-06-2018	0.87	0.53	1.00	-0.20
26-07-2018	0.09	0.58	-0.93	0.33
13-09-2018	-0.38	0.65	-0.50	0.00
25-10-2018	-0.60	0.62	-1.00	-0.30
13-12-2018	-0.70	0.45	-1.00	-0.30
24-01-2019	0.42	0.18	-0.84	-0.38
07-03-2019	0.46	0.12	-0.70	-0.30
10-04-2019	-0.38	0.09	-0.93	0.00
25-07-2019	-0.91	0.24	-1.00	-0.79
12-09-2019	-0.90	0.04	-1.00	-0.20
24-10-2019	-0.72	0.36	-1.00	-0.12

Note: This table presents the time series for vocal cues during the introductory statement ($Voice_t^{IS}$) and Q&A session ($Voice_t^{AN}$), as well as verbal cues during the introductory statement ($Language_t^{IS}$) and Q&A session ($Language_t^{AN}$).

Appendix B. Audio Processing

We convert all audio files to a 22050 Hz sample rate and mono channel. To extract MFCCs, we use the default options of the Librosa package: the Fast Fourier Transform window length is 2048, and the number of samples between successive frames is 512. We apply an orthonormal discrete cosine transformation and extract the first 100 MFCCs for each audio file. The audio files are neither cut nor preprocessed. Figure (B1) visualizes the MFCCs for a happy vocal emotion with normal intensity from RAVDESS:

Figure B1: **Visual Illustration of MFCCs**



Appendix C. Construction of a synchronized Audio-Language Data Set

For our analysis, we introduce a novel dataset that synchronizes audio and textual data of Mario Draghi's voice and language during the Q&A session. We download all audio data from the ECB Webcasts and convert them into WAV files. We exclude the introductory statements, moderator's interventions, journalists' questions, and the vice president's answers. The audio data for the introductory statement is processed independently of the Q&A session.

Typically, journalists at the ECB press conference ask two or three questions at once, including intermediate questions. Draghi usually responds to all questions in one sequence, even though the questions and answers often cover fundamentally different topics. Additionally, Draghi uses specific questions to share information about discussions during the ECB governing council meetings. To identify all individual answers within Draghi's contributions, we utilize a distinctive feature of the ECB press conference transcripts: the ECB staff separates the president's answers into distinct paragraphs. We adhere to this structure and edit the audio files to ensure synchronization with the text, maintaining the integrity of the message. Figure (C1) illustrates our approach:

Figure C1: **Illustration of Data Set Construction and Synchronization**

We apply this identification strategy to all press conferences held between May 2012 and October 2019. By manually screening all press conferences, we identify interjections by journalists that the ECB staff did not account for. In such cases, we remove the interjections and split the response into two separate answers: one before the interjection and one after. Our dataset comprises 71 press conferences, resulting in 2,336 individual answers as both audio and textual data.

Appendix D. FCN Model Structure Information

- **Convolutional Layers:** Convolutional layers perform a linear operation that involves multiplying a set of weights with the input data. This input can be raw audio data represented as an image or a feature map output from a previous convolutional layer. In our model framework, we utilize three convolutional layers. Each layer applies filters to the input data to detect various features. The first layer contains 64 filters with a kernel size of (11, 7), meaning it uses 64 sets of weights with a window size of 11x7 to process the input. The second layer also has 64 filters but with a kernel size of (7, 11). The third layer's number of filters corresponds to the number of emotion classes in our classification task, which is 6. These layers help the network learn to recognize patterns and features associated with different vocal emotions.

- **Activation Function:** Activation functions are used to propagate the output of one layer's nodes forward to the next layer. In our model, we use rectified linear units (ReLU), which are the most commonly used activation function in convolutional neural networks. The ReLU function outputs zero for any input value below zero, and for any input above zero, it outputs the input value itself, thus establishing a linear relationship:

  $$f(x) = \max(0, x) \tag{D.1}$$

  This non-linear transformation allows the network to learn complex patterns. For the final classification layer, we use a Softmax activation function, which is recommended for neural networks solving classification problems. Softmax converts the output into a probability distribution over the possible classes, ensuring the sum of the probabilities is one.

- **Dropout Layer:** The Dropout layer helps prevent overfitting by randomly setting a fraction of input units to zero during each training step. This process forces the network to learn more robust features by not relying too heavily on any single neuron. In our model, we use a dropout rate of 20%, meaning 20% of the input units are set to zero at each training step.
- **GlobalAveragePooling Layer:** The GlobalAveragePooling layer calculates the average output of each feature map from the previous layer. This operation reduces the data dimensions, preparing the model for the final classification step using a Softmax activation function. The GlobalAveragePooling layer extracts a single value from each filter, corresponding to the average of all filter weights. This approach allows for the analysis of non-fixed-length audio files, ensuring the model can handle varying input sizes effectively (García-Ordás et al., 2021).
- **Model Compilation:** We use "Adam" as the optimizer for model compilation, which is a stochastic gradient descent method known for its efficiency and adaptive learning rate. For model evaluation, we focus on accuracy, defined as:

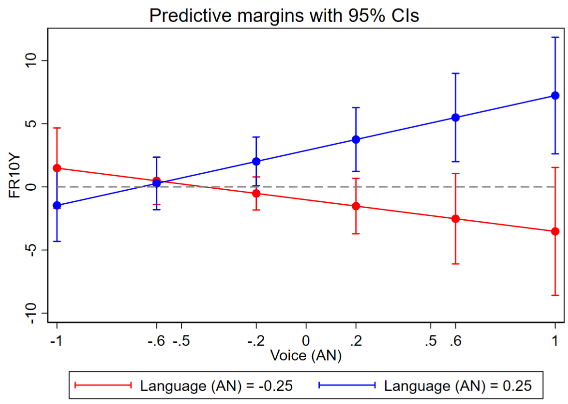
$$Accuracy = \frac{1}{N} \sum^{N} i = 1 \mathbb{1}(y_{i,\text{Pred}} = y_{i,\text{True}}) \tag{D.2}$$

N is the number of observations in our validation set and $\mathbb{1}(y_{i,Pred} = y_{i,True})$ is an indicator variable that equals 1 if our classification for observation i is correct, and 0 otherwise. We evaluate accuracy metrics exclusively on data that are not part of the training set. Our model uses a batch size of 80 and runs the training process for 2500 epochs. To avoid overfitting and reduce training time, we include an EarlyStopping callback with a patience of 100 epochs. If the model does not improve in out-of-sample accuracy after 100 epochs, the training stops, and the model with the highest accuracy is saved.

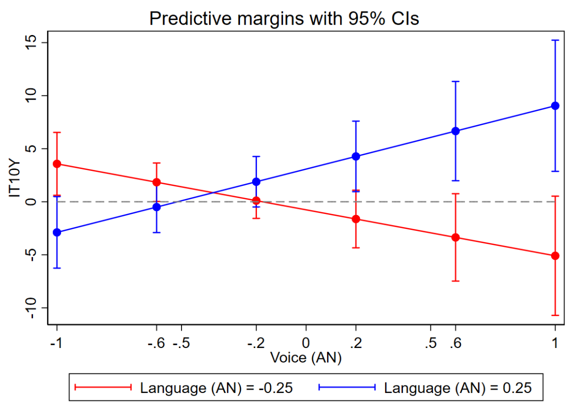
Appendix E. Additional Results

Appendix E.1. Additional Results for Yields and Spreads

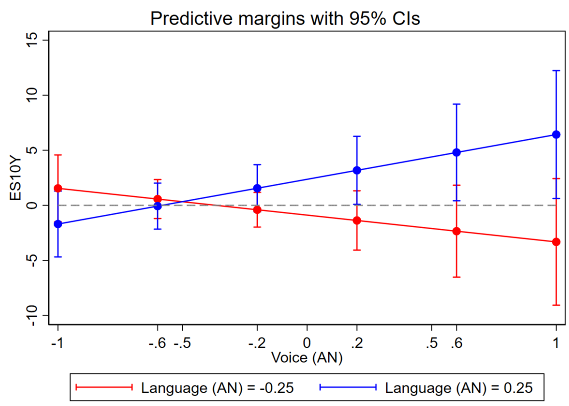
Figure E1: Marginal Effect of Vocal Emotions Given Language on Ten-Year Yields



(a) Ten-Year French Yield



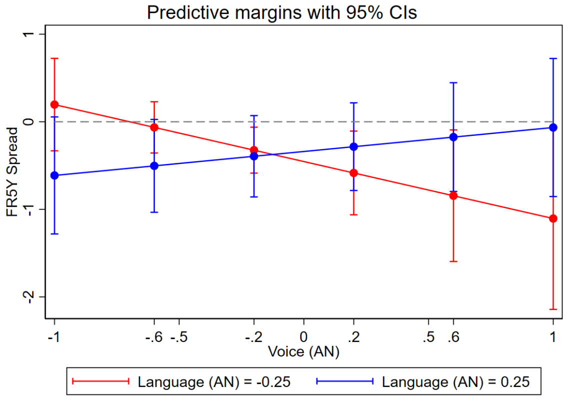
(b) Ten-Year Italian Yield



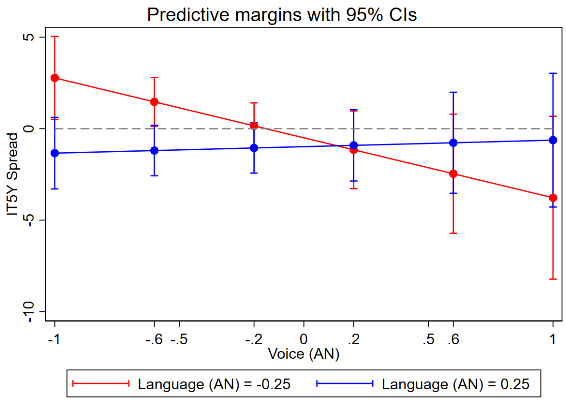
(c) Ten-Year Spanish Yield

Note: These plots visualize the marginal effect of a change in $Voice_t^{AN}$ given a specific level of $Language_t^{AN}$ on the yields of ten-year government bonds from a) France, b) Italy, and c) Spain. We report the change in yields in basis points. We illustrate the marginal effect of $Voice_t^{AN}$ for $Language_t^{AN} \in \{-0.25,\ 0.25\}$. We report the estimations in table (3).

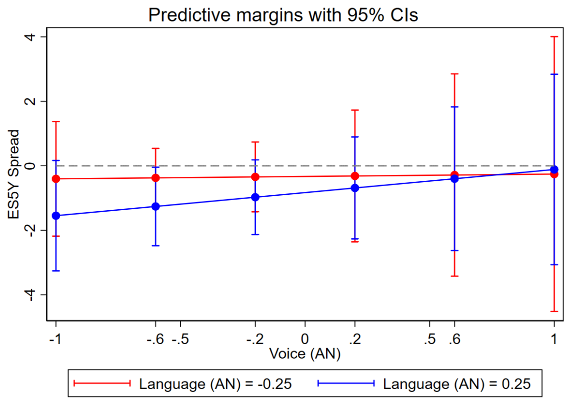
Figure E2: Marginal Effect of Vocal Emotions Given Language on Five-Year Spreads



(a) Five-Year French Spread



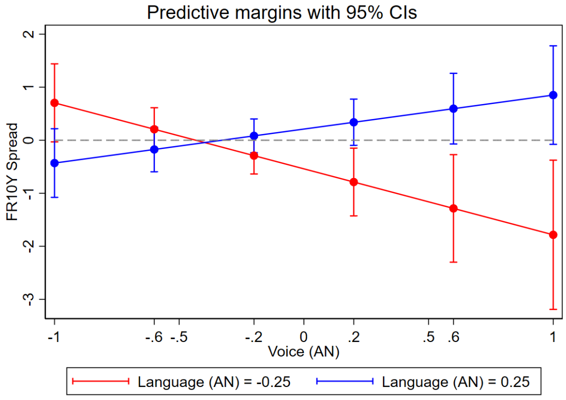
(b) Five-Year Italian Spread



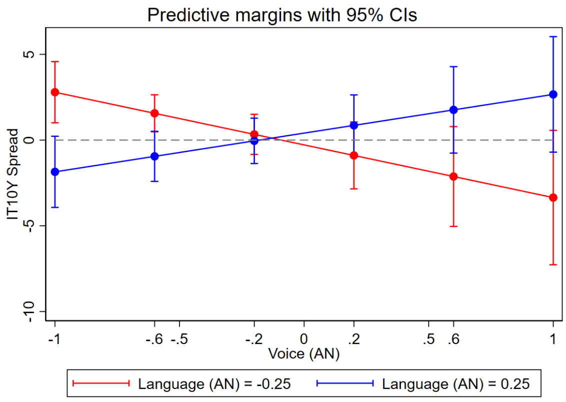
(c) Five-Year Spanish Spread

Note: These plots visualize the marginal effect of a change in $Voice_t^{AN}$ given a specific level of $Language_t^{AN}$ on the spreads of five-year government bonds from a) France, b) Italy, and c) Spain. We report the change in yields in basis points. We illustrate the marginal effect of $Voice_t^{AN}$ for $Language_t^{AN} \in \{-0.25,\ 0.25\}$.

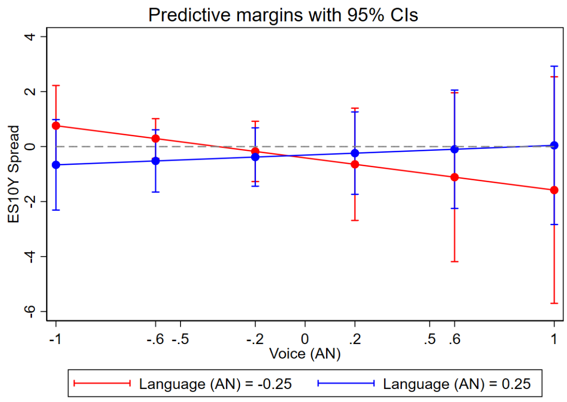
Figure E3: Marginal Effect of Vocal Emotions Given Language on Ten-Year Spreads



(a) Ten-Year French Spreads



(b) Ten-Year Italian Spreads



(c) Ten-Year Spanish Spreads

Note: These plots visualize the marginal effect of a change in $Voice_t^{AN}$ given a specific level of $Language_t^{AN}$ on the yields of ten-year government bonds from a) France, b) Italy, and c) Spain. We report the change in yields in basis points. We illustrate the marginal effect of $Voice_t^{AN}$ for $Language_t^{AN} \in \{-0.25,\ 0.25\}$.

Appendix E.2. Robustness Check: Likert Scale Definition for Vocal Emotions

To ensure that our conclusions are robust with a more qualitative definition of vocal emotions, we define $Voice_t^{IS}$ and $Voice_t^{AN}$ as follows:

$$Voice_t = \begin{cases} +2 & \text{if } \frac{Positive_t - Negative_t}{Positive_t + Negative_t} \in [+1,\ +0.6) \\ +1 & \text{if } \frac{Positive_t - Negative_t}{Positive_t + Negative_t} \in [+0.6,\ +0.2) \\ 0 & \text{if } \frac{Positive_t - Negative_t}{Positive_t + Negative_t} \in [+0.2,\ -0.2] \\ -1 & \text{if } \frac{Positive_t - Negative_t}{Positive_t + Negative_t} \in (-0.2,\ -0.6] \\ -2 & \text{if } \frac{Positive_t - Negative_t}{Positive_t + Negative_t} \in (-0.6,\ -1] \end{cases} \tag{E.1}$$

Our conclusions remain consistent even when using a qualitative classification of vocal emotions instead of a precise quantitative measurement.

Table E.1: **Robustness Check: Likert Scale Vocal Emotion Regression**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Dependent Variable	DE1Y	DE2Y	DE5Y	DE10Y	FR2Y	FR5Y	FR10Y	IT2Y	IT5Y	IT10Y	ES2Y	ES5Y	ES10Y
$Voice_t^{AN} \times Language_t^{AN}$	2.79***	3.32***	4.02***	4.27***	3.31***	4.84***	5.87***	5.71***	7.38***	9.01***	3.38***	4.98***	5.88***
	(0.63)	(0.84)	(1.04)	(1.32)	(0.90)	(1.14)	(1.42)	(1.14)	(1.66)	(1.91)	(1.12)	(1.44)	(1.72)
$Voice_t^{AN}$	-0.05	-0.08	0.22	0.56	-0.09	0.11	0.48	-0.50	-0.32	0.52	0.06	0.21	0.35
	(0.30)	(0.39)	(0.54)	(0.54)	(0.37)	(0.58)	(0.65)	(0.38)	(0.51)	(0.66)	(0.35)	(0.57)	(0.68)
$Language_t^{AN}$	4.41***	4.93***	5.81***	5.79**	4.64***	6.00***	7.12***	3.69**	4.94*	6.94**	4.12**	5.19**	6.23**
	(1.29)	(1.50)	(1.69)	(2.35)	(1.69)	(1.93)	(2.51)	(1.40)	(2.53)	(3.24)	(1.73)	(2.20)	(2.99)
$Voice_t^{IS}$	0.09	0.25	0.30	0.18	0.24	0.35	0.25	0.24	0.39	0.17	0.30	-0.10	0.19
	(0.32)	(0.44)	(0.62)	(0.66)	(0.41)	(0.68)	(0.79)	(0.49)	(0.63)	(0.80)	(0.38)	(0.65)	(0.79)
$Language_t^{IS}$	-1.13	-1.53	-2.95	-3.27	-1.10	-1.42	-2.00	1.66	1.70	0.44	-1.11	-0.36	-0.22
	(1.01)	(1.54)	(2.09)	(2.09)	(1.31)	(2.23)	(2.25)	(1.35)	(1.94)	(2.23)	(1.21)	(1.53)	(2.24)
$\Delta Inflation_t^{NCY,Positive}$	9.14***	12.06**	12.14**	5.48	11.47***	12.78**	9.55	13.46**	12.76**	12.49*	8.36**	11.80**	9.19
	(3.15)	(4.74)	(5.69)	(5.35)	(4.06)	(5.63)	(6.39)	(5.33)	(5.91)	(6.87)	(4.09)	(4.69)	(6.37)
$\Delta Inflation_t^{NCY,Negative}$	8.05	8.47	8.94	8.50	7.16	9.72	7.62	14.91	3.79	6.60	-1.44	1.94	3.36
	(5.31)	(6.44)	(8.87)	(8.55)	(6.12)	(8.87)	(9.01)	(11.09)	(10.01)	(10.47)	(6.26)	(9.16)	(8.71)
$\Delta RGDP_t^{NCY,Positive}$	-8.24**	-10.89**	-14.40***	-9.93*	-10.97***	-15.02***	-12.73**	-17.29***	-18.44***	-19.08***	-10.42**	-14.64***	-15.06**
	(3.20)	(4.19)	(5.08)	(5.42)	(3.87)	(5.10)	(5.86)	(4.82)	(5.71)	(6.47)	(4.14)	(4.90)	(5.99)
$\Delta RGDP_t^{NCY,Negative}$	1.22	-1.92	-8.86	-6.01	1.43	-5.47	-2.02	2.39	-0.91	-2.53	-2.97	3.00	-2.60
	(3.51)	(4.27)	(6.23)	(6.09)	(4.25)	(6.09)	(6.12)	(6.58)	(8.64)	(9.13)	(5.67)	(7.25)	(7.73)
Constant	0.57	0.97	2.22	2.45	0.61	1.03	1.62	-0.77	-0.32	1.04	0.60	0.42	0.91
	(0.68)	(0.98)	(1.35)	(1.36)	(0.86)	(1.43)	(1.47)	(1.03)	(1.37)	(1.56)	(0.95)	(1.27)	(1.54)
R^2	0.476	0.429	0.409	0.335	0.423	0.388	0.349	0.477	0.438	0.451	0.330	0.323	0.311
Obs	48	48	48	48	48	48	48	48	48	48	48	48	48

Note: The results are based on equation (1) using the change in the yield of government bonds of Germany, France, Italy, and Spain for August 2013 until October 2019 as the dependent variable (equation 4). The dependent variable is calculated as the difference in the median price in a narrow time window before the ECB press conference and a narrow time window afterward. We report the change in yields in basis points. $Voice_t^{IS}$ and $Voice_t^{AN}$ is Draghi's vocal sentiment during the introductory statement and Q&A session, respectively. We use a Likert scale definition for the vocal variables. Analogously, $Language_t^{IS}$ and $Language_t^{AN}$ measure the sentiment of Draghi's verbal cues. Furthermore, $\Delta Inflation^{NCY,Positive}$ ($\Delta RGDP^{NCY,Positive}$) controls for the change in the NCY forecast for inflation (real GDP) from the ECB/Eurosystem staff projections. Analogously, $\Delta Inflation^{NCY,Negative}$ ($\Delta RGDP^{NCY,Negative}$) controls for the effects of negative changes. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$