

Voice Beyond Words: Evidence That Managerial Tone Predicts Returns When Text Does Not

January 2026

Abstract

This research brief examines the incremental predictive power of managerial vocal cues in earnings calls, particularly in scenarios where textual sentiment is neutral. Building on the methodology of Ewertz (2025), we analyze a large dataset of Russell 3000 earnings calls and find that paralinguistic features, such as assertiveness, arousal, and nervousness, contain significant economic information. Our findings demonstrate that even when language is devoid of strong sentiment, the acoustic properties of executive speech can predict post-earnings announcement drift, with top-quintile portfolios based on vocal features generating alpha of 40–70 basis points over 10- to 30-day holding periods. This study provides robust evidence that non-verbal communication channels are a valuable and underutilized source of information for investment analysis.

1. Introduction

Investor communication is a multimodal phenomenon, where information is conveyed not only through the literal meaning of words but also through the nuances of their delivery. A growing body of literature in empirical finance has established that paralinguistic cues such as pitch, jitter, energy, and vocal tension contain economically meaningful information about managerial uncertainty, cognitive load, or confidence [1] [2]. Despite this, the majority of commercial earnings-call analytics continue to focus primarily on text-based sentiment analysis or keyword-based Natural Language Processing (NLP).

Speech Craft Analytics (SCA) extends this line of inquiry by extracting sentence-level acoustic, prosodic, and linguistic features from every Russell 3000 earnings call. This brief addresses a core question also raised by Ewertz (2025): Does voice add incremental predictive power when text is neutral? [2]. By isolating the impact of vocal cues from the confounding influence of textual sentiment, we aim to provide a clean and robust test of the informational content of managerial tone.

2. Data and Methodology

2.1. Data

The audio files for this analysis were sourced from S&P Global, covering all available Russell 3000 earnings calls. To ensure a focus on the most critical interactions, our analysis is restricted to the question-and-answer (Q&A) portion of the earnings calls, specifically focusing on the speech of the Chief Executive Officer (CEO) and Chief Financial Officer (CFO). To capture the full spectrum of vocal expression, including filler words, repetitions,

and disfluencies that are often scrubbed from commercial transcripts, SCA performs its own proprietary transcription of the calls. To reduce noise from short or incomplete utterances, sentences with fewer than four alpha tokens are disregarded.

2.2. Models

The vocal features used in this study are generated by proprietary SCA models. These models are based on speaker-specific baselines and are designed to summarize vocal delivery along economically interpretable dimensions using a wide range of acoustic features. The models were trained on a large corpus of earnings calls and calibrated on executive earnings calls to ensure their accuracy and robustness.

2.3. Experimental Design

Following the methodology of Ewertz (2025), we first condition on a mid-band of text sentiment to isolate the effect of vocal cues. This is achieved through the following steps:

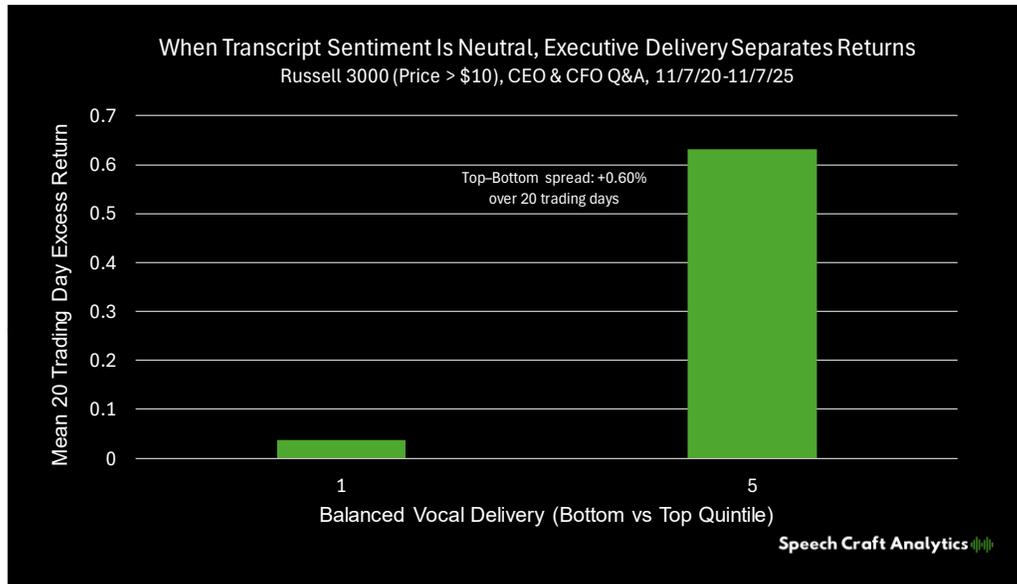
1. **Neutral-Text Conditioning:** We compute 90-day rolling percentiles for Sentiment Polarity, a measure of text sentiment. The sample is then restricted to instances where Sentiment Polarity falls within deciles 4–6, representing a ‘neutral’ textual tone. This conditioning effectively removes linguistic optimism or pessimism as a confounding variable, allowing for a focused analysis of the predictive power of voice alone. This process results in a universe of approximately 155,000 events.
2. **Ranking Voice Features:** Within this neutral-text subsample, we compute percentiles for each of SCA’s proprietary voice signals. These signals are then sorted into quintiles, with the top quintile (Q5) designated as the test portfolio for our analysis.
3. **Portfolio Construction and Returns:** For each holding period $H \in \{1, 5, 10, 15, 20, 25, 30\}$ days, we compute the H-day equal-weighted excess return against a price-screened, equal-weighted universe. Standard errors are clustered by event date to account for potential time-series dependencies, following the methodology of Petersen (2009).

3. Results

Our analysis reveals that even when the linguistic content of an earnings call is neutral, the way in which executives speak provides significant predictive power. As shown in the table below, features such as assertiveness, arousal calibration, and nervousness predict between 40 and 70 basis points of alpha over holding periods of 10 to 30 days, with highly significant t-statistics. These results are consistent with the findings of Ewertz (2025), confirming that conditioning on textual sentiment does not eliminate the predictive information contained in acoustic cues.

Long Only - Average Excess Returns by Holding Period 11/7/2020 - 11/7/2025
 Russell 3000 Universe where Sentiment Polarity is Neutral, CEO and CFO Q&A Comments, Price > \$10

	1 Day Holding Period Excess Return	5 Day Holding Period Excess Return	10 Day Holding Period Excess Return	15 Day Holding Period Excess Return	20 Day Holding Period Excess Return	25 Day Holding Period Excess Return	30 Day Holding Period Excess Return
Nervous_mgmt	-0.08%	0.15%	0.15%	0.30% *	0.37% *	0.43% *	0.49% *
Assertiveness_mgmt	0.01%	0.08%	0.10%	0.05%	0.18%	0.31%	0.40%
Balanced_mgmt	0.12% *	0.21% *	0.29% *	0.49% ***	0.63% ***	0.70% ***	0.65% ***
Valence_mgmt	0.01%	0.07%	0.06%	-0.03%	0.10%	0.26%	0.37%
Arousal_mgmt	-0.04%	0.23% *	0.30% *	0.46% **	0.53% ***	0.54% **	0.55% **
Entropy_mgmt	0.02%	0.10%	0.20%	0.34% *	0.35% *	0.46% **	0.34%



4. Discussion

The results of this study strongly indicate that vocal cues possess significant predictive value, even in the absence of a clear textual signal. The outperformance of management-level composite features suggests that these models are effectively capturing complex constructs such as conviction and cognitive load, which are not readily apparent from text alone. The fact that text-only features lose their predictive power under the neutral-text conditioning validates our experimental design and reinforces the conclusion that voice provides an independent channel of information.

5. Conclusion

This study provides compelling evidence that managerial voice contains material information for equity prediction, even when the language used is neutral and devoid of sentiment. Conditional on neutral text, the quality of vocal delivery remains a statistically and economically significant predictor of post-earnings announcement drift. These findings, which are consistent with those of Ewertz (2025), underscore the importance of incorporating paralinguistic analysis into any comprehensive evaluation of investor communications.

6. References

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

[2] Ewertz, S. (2025). Listen Closely: Measuring Vocal Tone in Corporate Disclosures, *Journal of Accounting Research*. August 2025.