Why Audio + Transcripts Don't Produce Behavioral Alpha

Asset managers increasingly explore the idea of extracting confidence, stress, or tone from earnings-call audio. Many already license the raw inputs: audio recordings and transcripts, and assume that with modern ML tools, building proprietary behavioral signals is a straightforward extension.

Access to audio and text is necessary, but it is nowhere near sufficient.

Across firms, internal build attempts repeatedly fail for the same fundamental reasons.

1. Earnings-Call Audio Is Not a Research Dataset

Earnings-call recordings are operational artifacts, not clean laboratory samples. They vary widely across companies, quarters, and telephony setups. Calls include multiple speakers, crosstalk, line noise, sirens in the background, paper rustling, and technical drift over time.

Small inconsistencies in segmentation, speaker attribution, or audio quality cascade into large differences in estimated behavior. Without dedicated audio-engineering infrastructure, the signal is unstable from the start.

2. Behavioral Voice Signals Require Domain-Specific Feature Design

Executive speech during earnings calls is cautious, strategic, and highly constrained. The cues that matter for confidence, uncertainty, or stress are subtle and context dependent.

General-purpose speech features and off-the-shelf models cannot detect the micro-variations in delivery that matter for financial interpretation. Behavioral voice analytics requires domain-specific design tuned to managerial communication under information asymmetry—not generic audio processing.

3. Domain Shift Breaks Standard Speech Models

Models built on everyday speech, broadcast speech, or expressive emotional datasets do not transfer to earnings calls. Executive vocal behavior follows a completely different statistical structure. Models must understand financial lexicon and be trained on executive speech cadence.

Teams often discover that internal models respond less to managerial effect and more to irrelevant attributes such as line clarity, room acoustics, or who speaks faster during the prepared remarks.

This is a structural problem, not a tuning problem.

© 2025 Speech Craft Analytics Inc. All Rights Reserved.

4. Cross-Sectional Comparability Is the Core Barrier

Even if a team extracts features and builds a model, the most difficult requirement remains:

Signals must be comparable across thousands of speakers, companies, sectors, and years.

Achieving this requires:

- rigorous speaker-level normalization
- consistent handling of scripted vs. unscripted speech
- corrections for technological drift in call infrastructure
- stable, longitudinal treatment of speaker baselines
- harmonization across heterogeneous environments

Most internal prototypes collapse here, producing signals dominated by operational artifacts rather than managerial behavior.

5. Investment Relevance Requires Deep, Multi-Year Validation

Producing a behavioral signal is not the same as producing alpha.

To make the signal investable, firms must validate it across multiple market cycles, different return horizons, sector structures, and alternative text-based controls, while continuously monitoring for drift and recalibrating models.

This requires infrastructure, time, and institutional process that few teams can sustain internally.

Conclusion

SCA Exists Because Internal Builds Don’t Work

Access to audio is not the hurdle.

Technology is not the hurdle.

Talent is not the hurdle.

The hurdle is that earnings-call speech is a complex, drifting, domain-specific signal that requires dedicated infrastructure, domain research, and continual recalibration.

© 2025 Speech Craft Analytics Inc. All Rights Reserved.

SCA provides what internal builds cannot deliver:

- a stable, investment-grade behavioral signal
- validated across years
- engineered specifically for financial communication
- maintained continuously
- and deployable immediately

Internal builds fail because they underestimate the complexity. Asset managers use SCA because we have already solved it.

© 2025 Speech Craft Analytics Inc. All Rights Reserved.