Learning from machines: Can CEO vocal cues help to predict future firm performance?

Zihao Liu
Tilburg University
liuzihao.connor@gmail.com

Abstract

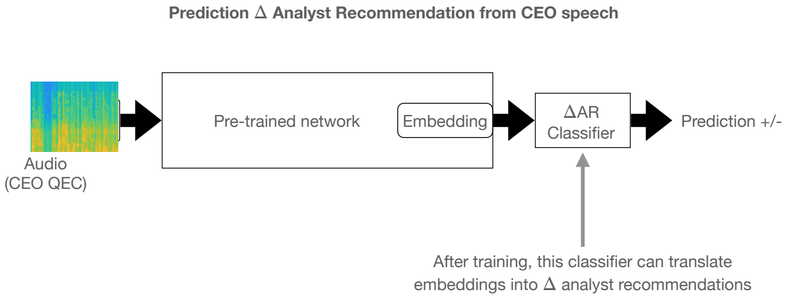
I apply tailored deep learning models on CEO voiceprints of earnings conference calls to predict the firm's future performance as measured by analyst recommendation consensus changes, unexpected earnings, and cumulative abnormal returns. The out-of-sample evaluations of the models consistently outperform established benchmark models using only textual information and firm characteristics by approximately 13% on average, achieving a prediction accuracy ranging from 54% to 65%. In other words, how firm information is communicated in addition to the content can affect the market perception of a firm. This study adds new evidence to audio recordings of conference calls containing valuable information about a firm's fundamentals, incremental to qualitative "soft" information conveyed by textual content, and quantitative earnings information. To achieve some level of explainability of the learned voiceprints' nuances, I employ a tailored vocal emotion classifier, contributing to improving and refining vocal sentiment analysis in the existing literature.

Research questions

1. **To what extent do vocal characteristics of a CEO predict a firm's future financial performance?**
2. Which vocal characteristics are important for predictions?
3. What are the economic mechanisms of the vocal characteristics?
    1. Manager quality traits
    2. Private information
    3. Behavioral biases

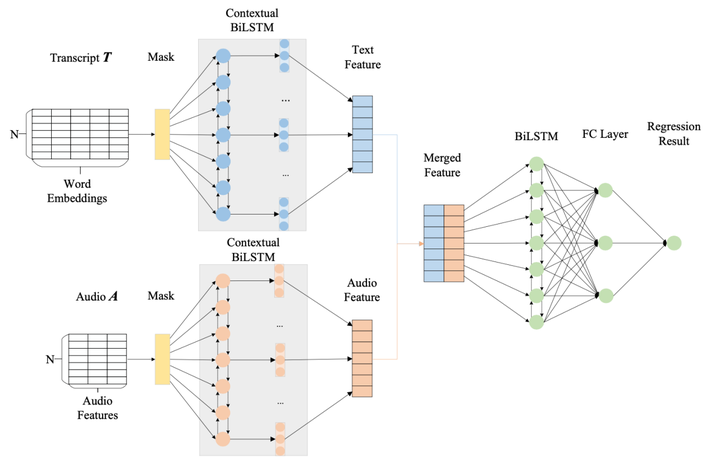
I: Global Measure on CEO Vocal Cues

- **Input:** CEO voice print matrices
- **Output:** Analyst recommendation consensus changes after calls
- Using transfer learning techniques to repurpose related pre-trained deep learning models based on CNNs such as SpeechVGG
- The embedding is the numerical representation of CEO vocal cues



II: Multimodal Approach

- I incorporate the audio feature together with the textual feature and financial feature by using a multimodal approach to predict firm performance out-of-sample and compare results with the benchmark models without audio feature.



Vocal model prediction comparisons

Model Architecture	Δ REC (in a month)		improvements
	SpeechVGG	**65.1%**	29.2%
	CI (99%)	[62.7%, 67.4%]	
	2CNN(Max)	**57.3%**	13.7%
	CI (99%)	[55.3%, 59.4%]	
	UE(in a quarter)		
	SpeechVGG	**58.4%**	13.8%
	CI (99%)	[56.8%, 60.0%]	
	2CNN(Max)	**54.8%**	6.8%
	CI (99%)	[53.3%, 56.4%]	
	CAR(0,2)		
	SpeechVGG	**61.7%**	18.7%
	CI (99%)	[59.6%, 63.8%]	
	2CNN(Max)	**52.5%**	1.0%
	CI (99%)	[50.8%, 54.3%]	
	CAR(half-year)		
	SpeechVGG	**65.7%**	28.6%
	CI (99%)	[63.7%, 67.7%]	
	2CNN(Max)	**54.9%**	7.4%
	CI (99%)	[53.0%, 56.7%]	

1. The prediction results are based on binary dependent variables (50-50 test)
2. The data (train/test split) is reshuffled 100 times
3. All of the average model performance is better than textual model benchmarks
4. The transfer-learning model architecture (SpeechVGG) works much better

Control for firm characteristics

Model Architecture	Δ REC (in a month)		improvements
	SpeechVGG	**62.4%**	16.4%
	CI (99%)	[60.2%, 64.5%]	
	2CNN(Max)	**57.6%**	7.5%
	CI (99%)	[55.8%, 59.4%]	
	UE(in a quarter)		
	SpeechVGG	**60.4%**	4.9%
	CI (99%)	[58.6%, 62.2%]	
	2CNN(Max)	**58.1%**	0.9%
	CI (99%)	[56.3%, 59.9%]	
	CAR(0,2)		
	SpeechVGG	**62.5%**	16.8%
	CI (99%)	[60.6%, 64.3%]	
	2CNN(Max)	**56.3%**	5.2%
	CI (99%)	[54.5%, 58.2%]	
	CAR(half-year)		
	SpeechVGG	**63.9%**	19.7%
	CI (99%)	[61.6%, 66.1%]	
	2CNN(Max)	**54.2%**	1.5%
	CI (99%)	[52.6%, 55.9%]	

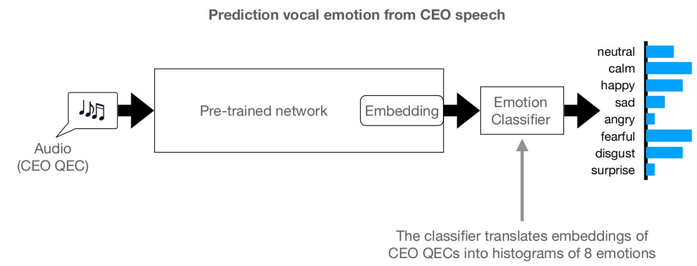
1. The additional control variables include size, B/M ratio, return volatility, profitability, and momentum
2. The overall accuracies increased slightly on average
3. The improvements are weaker but still economically meaningful.

Reverse-engineering to interpret

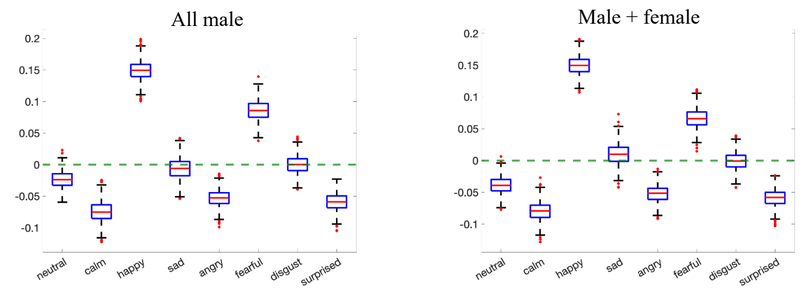
1. One of the key innovations of this paper is that the audio features are not only subjective to emotions (transitory information) since I train vocal embeddings directly on firm performance measures.
2. The global measure can incorporate both transitory and more permanent information as long as they relate to firm performance.

Interpret embeddings of extracted audio features

- Using the extracted vocal embeddings to classify emotions



Compare emotion scores between positive and negative speeches



1. The positive CEO speeches are happier and more fearful than the negative ones
2. While happy emotion can relate to the positiveness of a CEO's affective state, the fearful emotion needs further investigation

Conclusion

1. Deep learning models appears successfully extracts vocal cues for both short-term and long-term firm performance predictions
2. The emotions of CEO speeches can partially explain the extracted CEO vocal features
3. The economic mechanisms can be further investigated by using the global measure of CEO vocal cues
4. I apply more scientific and transparent methodologies based on customized DL algorithms instead of commercial software or pre-trained algorithms on small samples. My approach offers several advantages in terms of flexibility, room for further development, and implementations.