NOT FOR PUBLIC RELEASE

The Tone of Voice Provides a Novel Source of Alpha

Gerwin Schalk[1]

[1] Helios Life Enterprises, Albany, New York
* Correspondence: info@helioslife.enterprises

Abstract

Markets are influenced in important ways by earnings conference calls. For many years, investors have been guiding their decisions based on information derived from *what* words an executive is saying. While we know from decades of research and also from personal experience that the tone of the voice, i.e., *how* words are being spoken, holds important and independent information, quantitative information about voice tone has not been widely available and thus not been systematically incorporated in decision-making. In this White Paper, we describe the first widely available data product that systematically assesses the tone of the voice of an executive during earnings conference calls to produce novel and meaningful sources of quantitative information. We provide the principal reasons that voice tone is a highly desirable source of important and independent information about an executive's assessment of the firm, summarize the rationale and supporting methods for turning voice recordings of earnings calls into quantitative information, and describe the value of these data in producing novel, actionable, and predictive sources of alpha.

Keywords: voice, earnings call, voice tone, equity variables, merger and acquisition, cumulative abnormal return

"The effect of emotions upon the voice is recognized by all people. Even the most primitive can recognize the tones of love and fear and anger; and this knowledge is shared by the animals. The dog, the horse, and many other animals can understand the meaning of the human voice. The language of the tones is the oldest and most universal of all our means of communication."
— Smiley Blanton (Blanton, 1915)

1. Introduction

1.1. Evaluation of Earnings Calls

Executives distribute substantial amounts of information in earnings conference calls. Traditional ways to extract this information have focused primarily on fundamental data such as company performance (Campbell and Shiller, 1988, Bacidore et al., 1997, Penman, 1992, Heaton and Lucas, 1999) or text-based sentiment (Deng et al., 2011, Bing et al., 2014, Bharathi and Geetha, 2017). Hence, the reaction of the market, which depends in substantial ways on models that incorporate this information, rests in important ways on *what* was said by the executive.

1.2. Information in Communication

The seminal work by Ekman and Friesen (1969) described that humans communicate through different channels such as words, body posture, facial expression, and the tone of the voice (Rosenthal and DePaulo, 1979), and that some of these channels can be controlled better than others by an individual. In particular, several studies indicated that the tone of the voice gives away specific information that is not present in the verbal content (Weitz, 1972, Bugental et al., 1980, Bugental and Love, 1975, Bugental et al., 1976), can be less well controlled than facial expressions (Zuckerman et al., 1981), and is a better indicator of deception than facial expressions (Zuckerman et al., 1982). Indeed, it is becoming increasingly accepted that facial expressions are not reliable passive indicators of emotions but rather are actively used by humans to guide social interactions (Fridlund, 1997). Thus, it is not surprising that communication by voice only (i.e., without access to facial expressions) enhances the accuracy of the detection of different affective states (Kraus, 2017).

In sum, voice is not only much more ubiquitous than other types of information such as video, it is also best suited to detect the emotional state of the executive.

1.3. Voice Tone Gives Independent Information

Most people have experienced social situations in which the tone of the voice provides information that is different from the words a person is saying. This personal experience is strongly backed up by the scientific literature. For example, the social psychology literature has shown that the tone of the words, i.e., *how* the words are being spoken, contains about 40% of independent information in a message (Mehrabian and Wiener, 1967, Mehrabian and Ferris, 1967, Mehrabian et al., 1971, Walker and Trimboli, 1989, Caffi and Janney, 1994). This substantial independence of information in *what* vs. *how* words are spoken is further supported by neurolinguistics, which suggests that these different aspects of speech are even generated by different hemispheres in the brain.

The left hemisphere in the brain is concerned with the syntactic, semantic, and motoric aspects of speech perception and production (Leuthardt et al., 2007, Schalk et al., 2008, Brunner et al., 2009, Breshears et al., 2010, Roland et al., 2010, Pei, Leuthardt, Gaona, Brunner, Wolpaw and Schalk, 2011, Leuthardt et al., 2012, Potes et al., 2012, Kubanek et al., 2013, Sturm et al., 2014, Lotte et al., 2015, de Pesters et al., 2016, Taplin et al., 2016, Fedorenko et al., 2016, Brumberg et al., 2016). The information about these aspects in the brain is so detailed that, with appropriate detection and machine learning methods, it is possible to decode vowels, consonants, and even full words and sentences from brain signals alone (Pei, Barbour, Leuthardt and Schalk, 2011, Pei et al., 2012, Martin et al., 2014, Herff et al., 2015, Martin et al., 2016, Riès et al., 2017).

In marked contrast, the right hemisphere of the brain is also involved in speech perception (Chang et al., 2011, Swift et al., 2018) and production (Cogan et al., 2014), but is primarily specialized in affective components of language (Ross and Mesulam, 1979, Ross, 1981). With appropriate detection and machine learning methods, it is possible to decode different affective states from brain signals alone (Ethofer et al., 2009, Frühholz et al., 2012, Kim et al., 2013, Kragel and LaBar, 2016).

In summary, there is overwhelming evidence from the social psychology and neurolingustic literature that the tone of the voice gives information that is substantially different from that captured by the words that are being spoken.

1.4. Voice Tone is Valuable in the Context of Earnings Calls

To understand how and why the tone of the voice is particularly relevant and valuable in the context of earnings calls, it is helpful to understand how the emotions that underlie voice tone are being produced. While there are different theories of the emotion process, a popular and particularly applicable one is the appraisal theory (Roseman, 1984). According to that theory, voice tone is a response to events given a person's understanding of specific circumstances (Scherer et al., 2001).

In the context of earnings conference calls, this concept suggests that an executive exhibits specific tonal signatures that represent his/her understanding of company fundamentals. Thus, the tone of the voice may or may not coincide with the specific words they are saying, in particular if the company is under stress. In this case, the executive may be coached to produce positive statements, but will leak important and differing information in his/her tone, e.g., by sounding depressed or dismissive. For the same reason, this emotional information should predominantly be contained in answers to scrutinizing analyst questions, and not in the relatively scripted and rehearsed introductory remarks of an earnings call.

In summary, research in different fields strongly supports the notion that: 1) humans produce characteristic tonal signatures that reflect their understanding of company fundamentals; 2) the information in voice tone is substantially independent of that contained in words and can also not be well controlled/suppressed by the executive; and 3) specifically in particularly important situations (such as when the company is under stress), an executive's voice tone leaks information about the organization that is not available from any other source.

For all these reasons, systematically and quantitatively assessing tone from voice recordings in earnings conference calls should unlock an important and previously not readily available source of alpha.

1.5. Availability of Voice-Based Alpha

Helios produces the first and currently only widely available data products that deliver systematic analytics of an executive's voice tone in earnings conference calls. In this White Paper, we describe the rationale for and principles of voice analytics, the quantitative outputs that are supported by these analytics, and strong evidence that these data hold novel, substantial, and independent predictive information about different equity variables.

Important Comment: The purpose of this White Paper is to disclose relevant details about the generation and validation of our data, and to document its value in alpha generation. While the general style of this document is that of an academic research paper, we omitted specific details that would enable replication of these procedures. Also, the information in this White Paper is considered proprietary information for the purpose of Helios non-disclosure agreements.

2. Methods

2.1. Models of Human Emotions

To understand how to best extract features in voice recordings that reflect emotions, it is helpful to understand how emotions can be principally characterized.

Specific quantitative realities (such as the number of cars produced by a particular factory) can readily be measured with appropriate sensors (such as satellite images) and methods (such as computer vision). In contrast, emotions generally and voice tone specifically are qualitative and subjective characterizations, and so it is less clear how to best measure them.

Principally, there are categorical and dimensional ways to think about emotions. One important example of categorical models of emotions has been put forward by Ekman (1992) who argued that there are six basic emotions (anger, disgust, fear, happiness, sadness, and surprise), and that individuals can express each of them independently to varying degrees. While this model has been widely accepted, it is not clear how the subtle tonal cues likely expressed by an executive during a conference call (such as increased pausing or decreased rhythmicity) would map to these emotional categories.

Dimensional models appear to be better suited to the task at hand, because they make quantitative assessments of specific emotional states according to specific measureable dimensions. Two of the most common dimensions are arousal and valence (Banse and Scherer, 1996, Barrett, 1998). Arousal is usually measured from low to high whereas valence is usually measured from negative to positive. This model appears to be useful, because opposite emotions (such as happy and sad) map to opposite points in this two-dimensional arousal/valence space.

Based on these considerations, we extracted from the voice specific features that have been shown by prior acoustic and neurolinguistic research to be reflective of differing levels of valence or arousal. Generally, prior research has shown that arousal is usually well captured by acoustic measurements, while valence can often be better measured using semantic features.

2.2. Voice Features

Helios extracts 26 different features of the executive's voice from the audio recordings. These features include prosodic features, i.e., suprasegmental and nonverbal aspects of speech, such as, conversationally speaking, voice intonation, accent, speed, volume, and inflection (Scherer et al., 2003), as well as other acoustic and neurolinguistic qualities of the tone of the voice.

If multiple executives (e.g., the CEO and CFO) were giving the introduction or answering questions, these voice features were derived from the primary executive (usually the CEO) in the earnings call. Importantly, the procedure to extract any of these voice features is not informed in any way by any future information, or any financial target variable.

We reference these 26 features within and across earnings calls to account for differences in tonal characteristics across different executives. Similar to commonly available measurements for text-based sentiment, we make those 26 features available separately for the prepared remarks and the Q&A periods of an earnings call. Finally, we produce two different statistical moments (mean and standard deviation) from the many measurements of those features during each call, resulting in a total of 104 feature values (26 x 2 x 2) for each call. In essence, these feature values give a fingerprint of the characteristic tonal changes in voice of the executive during this call.

These data form the basis for two products: *Helios COMPREHEND* gives the 104 pre-processed voice features, and *Helios MERCURY* provides a quantitative assessment of the distress experienced by the executive during the earnings call.

For each product, our historical dataset covers 58580 audio recordings of earnings calls sourced from an enterprise-level data provider. They cover a period from 2005 to early 2021. The earnings calls are from a total of 1188 equities that represent the rebalanced Russell 1000. Thus, there is no survivorship bias in the data. Also, we did not exclude any samples from our dataset for any reason. Thus, we did not add to the minimal selection bias inherently imposed by the R1000.

3. *COMPREHEND:* Pre-processed Voice Features

3.1. Introduction

Helios COMPREHEND provides the 104 pre-processed voice features described above that together describe the emotional state of the executive during an earnings call. These features are predictive of different equity variables as described in the following sections. Because these data require further modeling/analyses to translate them into specific predictions, they are most useful to quantitative researchers.

In contrast to live data, which are delivered through our API, we deliver these historical data as a CSV file. Its contents consist of many rows — one row for each conference call. The format of each row is described in a data dictionary provided with the CSV file.

3.2. Evaluation Methods

The 104 raw voice features in Helios COMPREHEND hold information that allows for predictions of specific equity variables measured after the earnings call. Because these data are solely derived from the tone of the voice of executives during earnings calls, they do not include or are in other ways informed by well-known control variables (such as book-to-market ratio) that are readily available to others. The following sections describe how these predictions are generated, and the results implied by these predictions. (Inclusion of control variables in these predictions substantially further improves the results described below.)

3.2.1. Model Generation

We used support vector regression (Drucker et al., 1997, Smola and Schölkopf, 2004) to establish the relationship between an executive's voice features during the conference call with each considered equity variable (such as the earnings surprise), and applied this understanding to unseen data.

Specifically, for each equity variable, we trained a model on a set of earnings calls and tested it on the remaining earnings calls. We then repeated this process for all calls so that each time, different earnings calls became testing data. We then concatenated the predictions of equity variables across all these calls.

3.2.2. Model Evaluation

We evaluate the performance of our predictive models by calculating the Spearman correlation (r) between the predicted and actual equity variables.

3.2.3. Statistical Testing

To identify whether our procedure produces forward-looking predictions that are better than random, and to also ensure that there are no algorithmic biases that suggest spurious sources of information, we determined whether our r values are statistically significantly higher than those expected by chance. To do this, we applied a bootstrap randomization test (Efron and Tibshirani, 1993) in which we randomly scrambled the equity variables across conference calls 200 times, and computed one random Spearman's r value for each such iteration, which resulted in 200 measurements of randomized r values. We modeled these measurements using a Gaussian distribution (i.e., we calculated the r mean and standard deviation)[1]. We finally determined the probability p that the observed value of r was generated by the Gaussian model distribution of randomly generated r values. With this procedure, we kept all statistical properties of both the input voice features as well as the equity target variables, and only destroyed the temporal relationship between the two.

[1] We assessed the normality of these measurements using a Kolmogorov–Smirnoff (Massey Jr, 1951) test. This test determined that 93% of all distributions were considered Gaussian at the 0.05 level.

3.3. Voice Tone Predicts Equity Variables

The comparison of the actual and predicted equity variables suggests that the voice-based features available in *COMPREHEND* are informative of aspects of the reaction of the market (e.g., immediate volatility), reaction of the analysts (e.g., next quarter's earnings surprise), as well as the future profitability of the firm (e.g., Tobin's Q). We are well aware of the practical issues (such as actionability and/or *Post-Earnings Drift* (Bernard and Thomas, 1989) around short-term predictions and that, thus, corresponding information in voice tone may only be useful in certain circumstances.

However, our results strongly suggest that voice-based tone holds information about equity parameters that are measured at least a quarter after the earnings call. Presumably, an executive will be aware of specific company fundamentals and this information will leak in his/her voice tone. Because this information captured in voice tone is not reflected in fundamental data or text-based sentiment, it will not be systematically priced in by analysts. In line with this expectation, voice-based tone supports robust predictions of next quarter's earnings surprise (r = 0.069, p << 0.001) or other variables such as Tobin's Q (r = 0.168, p << 0.001).

Despite this encouraging finding, and despite the prior research discussed in Section 1.3 and in Mayew and Venkatachalam (2012), it is still possible that information based on voice tone may be shared with information from other sources, in particular important control variables or text-based sentiment. The following three sections forcefully argue that voice-based tone provides independent information relative to more conventional information, and that its inclusion improves predictions of earnings surprise beyond that provided by text-based sentiment derived from the executive's words during the same earnings calls.

3.3.1. Information from Voice Tone Provides Independent Information

Evaluating the independence of information from voice-based tone, we considered the following 8 control variables that have been used in previous studies (Collins and Kothari, 1989, Mayew and Venkatachalam, 2012):

1. Book-to-Market Ratio: book value of equity divided by market value of equity at end of quarter
2. Natural logarithm of the market value of equity
3. Unexpected Earnings: (actual earnings-mean analysts' earnings forecast)/stock price
4. Momentum: CAR(-127, -2)
5. Stock Price Volatility: standard deviation of daily stock returns over the period (-127, -2) relative to earnings call date
6. Raw Return: daily return between earnings conference call date and previous date
7. Year (fixed effect)
8. Month (fixed effect)

For text-based sentiment, we derived the number of positive and negative words spoken by the executive during the same earnings call using the list described in (Loughran and McDonald, 2011), expressed them as the fraction of total words, and calculated these fractions of positive and negative words separately for: 1) the whole transcript; 2) the prepared remarks; and 3) the Q&A period. We derived one measurement of text-based sentiment for each earnings call using the same methods (support-vector regression predicting earnings surprise, cross-validation) used for voice-based tone.

Finally, we determined Spearman's r between actual and predicted earnings surprise measurements in five ways:

1. voice tone alone (r = 0.069, as above)
2. voice tone after accounting for the 8 control variables using partial correlation
3. text-based sentiment alone
4. text-based sentiment after accounting for the 8 control variables
5. voice tone after accounting for the 8 control variables and text-based sentiment

The results of these analyses are shown in Fig. 1. They demonstrate that voice-based tone provides robust information about the earnings surprise (r = 0.038, p << 0.001) even after accounting for important control variables and text-based sentiment.

3.3.2. Voice Tone is Particularly Informative When it Goes Against Text-Based Sentiment

We know that what executives say will often concur with how they are saying it — presumably when their understanding of the firm coincides with their description of it. Thus, when the effect of text-based sentiment is removed, a fraction of earnings calls will have little or no distinctive voice-based information, thereby diminishing alpha when calculated across all calls as in the right-most bar in Fig. 1.

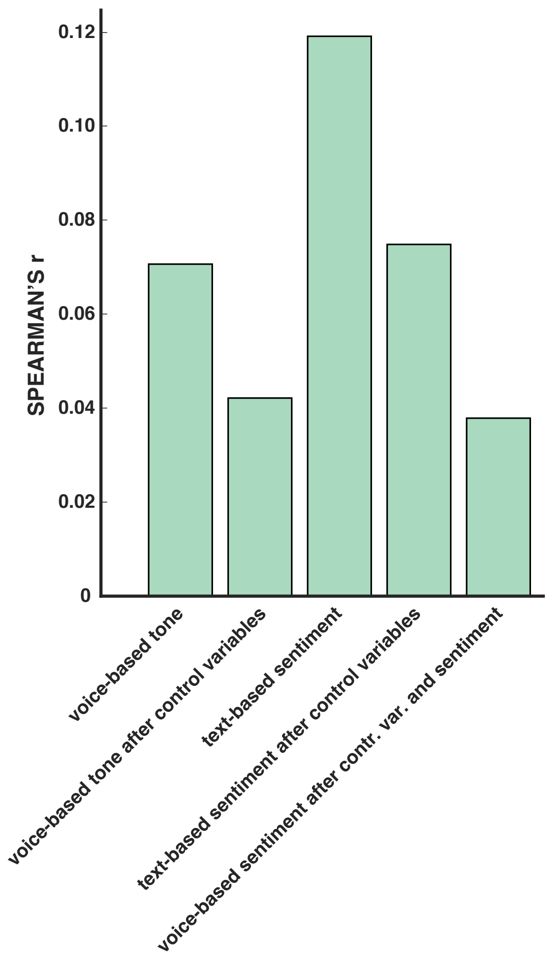

Figure 1. Voice-based tone gives independent information. The bars give r values of Spearman's direct/partial correlation of predictions of the earnings surprise for voice-based tone, text-based sentiment, as well as their contrasts to control variables.

Hence, identifying the cases in which the information in voice tone goes against that of text-based sentiment. Our additional analyses suggest that this is indeed the case. Specifically, we excluded different fractions of earnings calls in which voice-based information was the most similar to that provided by text-based sentiment, and calculated partial Spearman's r for voice-based tone against the control variables and text-based sentiment (as done above) for the remaining calls. When we excluded 25% or 50% of earnings calls, we achieved further improvements in our predictions ($r = 0.039$ and $r = 0.043$, respectively). Thus, this or similar procedures may be useful for identifying those cases in which voice-based tone provides unique information.

3.3.3. Voice-Based Tone Adds Information to Text-Based Sentiment

To produce the final piece of evidence that voice-based tone adds to information in text-based sentiment, we produced predictions of the earnings surprise from: 1) voice-based tone; 2) text-based sentiment; and 3) voice-based tone and text-based sentiment together. The results are shown in Fig. 2. They document that information from voice-based tone adds distinctive information to predictions of the earnings surprise based on text-based sentiment alone.

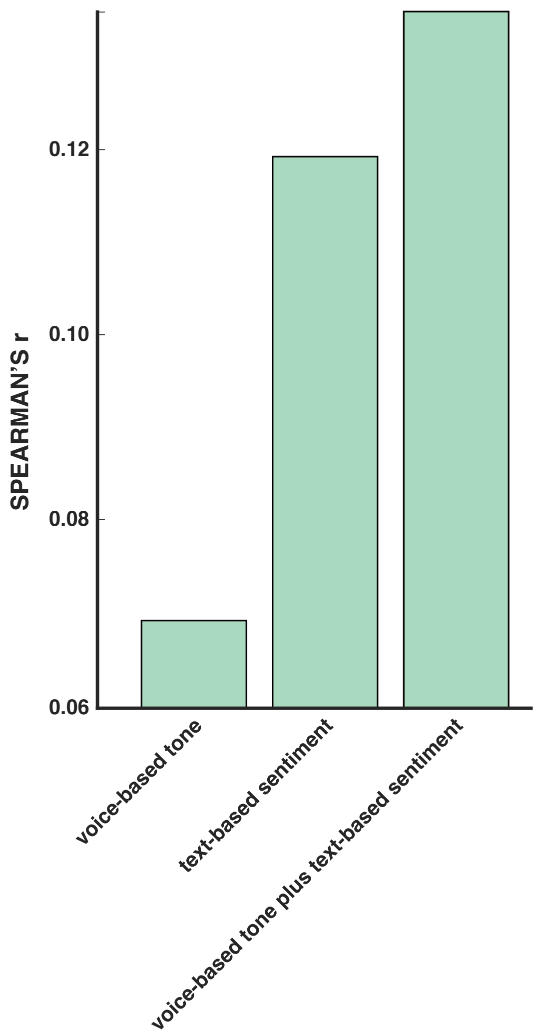

Figure 2. Voice-based tone adds information to that given by text-based sentiment. The bars give r values of Spearman's correlation of predictions of the earnings surprise for voice-based tone, text-based sentiment, as well as for both sources of information together.

3.4. Voice-Based Tone Provides Novel and Valuable Insights Into M&A Events

3.4.1. Mergers and Acquisitions

Mergers and acquisitions (M&As) play a key role in sculpting a firm's corporate structure, and, thus, are one of the most important transactions a firm can engage in. The term M&A stands for a transaction that takes place in the market for corporate control, where one firm uses cash or stock to acquire another firm's equity (Manne, 1965). Mergers and acquisitions have traditionally been an essential tool for growth of many organizations. For example, Alexandridis et al. (2010) provided evidence that M&A deals create value that leads to positive and statistically significant excessive returns for acquiring organizations.

M&A events have routinely attracted substantial attention in the media, such as Microsoft's purchase of Nokia's handset and services business for $7.2b in 2013, or Morgan Stanley's recent $13b purchase of E*Trade Financial in early 2020[2]. Many academic studies have evaluated the relationship between different characteristics of the firm or the merger/acquisition and the reaction of the marget (e.g., Travlos (1987), Schwert (2000), Gondhalekar et al. (2004), Moeller et al. (2004), Masulis et al. (2007), Rodrigues and Stegemoller (2014); Betton et al. (2008) for a detailed review), and have produced important insights into the M&A process. For instance, Travlos (1987) revealed significant differences in abnormal returns between common stock exchanges and cash offers (Masulis et al., 2007) in that acquirers with more anti-takeover provisions experience significantly lower announcement-period abnormal stock returns.

Research also shows that the explanatory power of models derived from these insights is rather limited (Routledge et al., 2017). This is unfortunate, because M&A events provide significant opportunities for investors as they often lead to a large and rapid change in market prices. Consequently, any improvement in the ability to predict which firms may be involved in upcoming M&A announcements and how the announcement will be evaluated by the market should prove to be highly profitable.

The following sections illustrate, using a large database of more than 10,000 of M&A announcements, that Helios COMPREHEND data provide clues about: 1) an upcoming merger announcement; 2) firm performance in the year following the announcement; 3) cumulative abnormal returns three months or more after the announcement; and suggest that this information should prove to be valuable in investments in the M&A space.

[2] https://www.washingtonpost.com/business/2020/02/20/morgann-stanley-etrade/

3.4.2. Database of M&A Events

The source of the data in the following subsections are 10,018 M&A events from 2005 to 2020 in Thomson Reuters' SDC Platinum database, as well as associated measures of future firm performance from COMPUSTAT and measurements of abnormal returns from CRSP. We restricted our sample to merger announcements where the acquiring firm was public and part of Helios' data universe, and where the deal value was larger than $ 1 million. Additionally, we focused on only those deals for which the acquiring firm owned less than 50% of the target's shares before the announcement.

We evaluated future firm performance given eight different measurements: net profit, price/operating ratio, price/earnings ratio, return on asset (ROA), return on equity (ROE), Tobin's Q, cash/liability ratio, and debt/asset ratio. We calculated these variables for the fiscal year following the announcement.

We also evaluated cumulative abnormal returns at 1, 2, 3, 4, 5, 6, 30, 60, 90, 180, and 360 days after the M&A announcement. These CAR measurements were calculated using total returns (i.e., the adjusted close), and hence include splits and dividends. Thus, even if executives only leak moderate information about future abnormal returns in their voice tone, this information would be highly actionable and should support profitable trading.

3.4.3. Voice Tone Predicts Whether or not a Firm May Announce a Merger

We first determined whether voice tone can be used to stratify firms into those that may or may not announce M&A events, and thereby hypothesized that the tone of the voice of the executive provides clues that are only found around M&A events.

We evaluated this hypothesis by first identifying all 25,934 earnings calls between 720 days prior to and 480 days following each of the 10,018 M&A announcements. We then asked whether voice tone can differentiate the period immediately prior to and after an M&A

announcement (360 days prior to and 180 days; M&A period) from other times in which no announcement was made (all other times; no M&A period). To do this, we labeled all 15,136 earnings calls in the M&A period with 1, and the remaining 10,798 earnings calls with 0, and randomly subsampled 10,798 earnings calls from the pool of 15,136 calls. Thus, this procedure provided an equal number of earnings calls during the M&A and no M&A periods. Hence in absence of any information in voice tone, random accuracy in predicting either of these two periods would be 50%.

We then fed the 104 voice-based features and M&A/no M&A labels for the 21,596 earnings calls to a support-vector classifier (10-fold cross validation) to determine to what extent the voice-based features can predict the correct 0/1 label and, thereby, the correct period. The results show that this procedure correctly predicted 51.8% of earnings calls and that actual and predicted labels were significantly correlated with each other (r=0.0365, p<<0.001, Spearman's correlation). Moreover, when we averaged the predictions in each quarter prior to and following an M&A announcement, the results (red trace in Fig. 3 and corresponding standard error) clearly reveal that an executive begins to leak information about the upcoming M&A announcement approximately 5 quarters prior to the announcement. We hypothesize that these vocal clues reflect the executive's increasingly emotional reaction (such as anxiety) leading up to the official announcement.

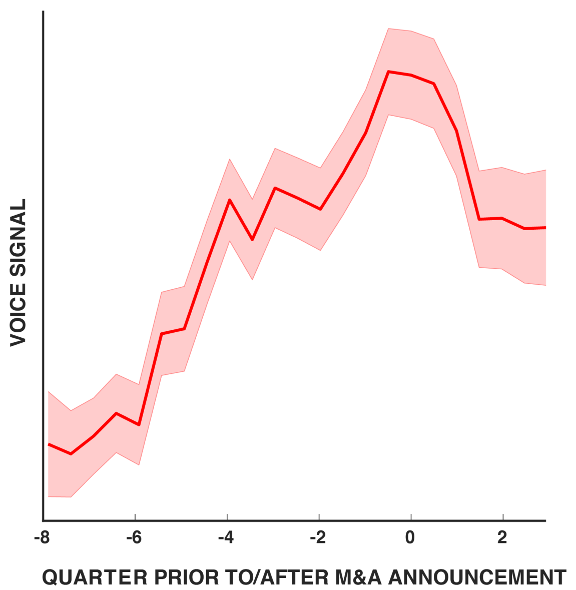

Figure 3. Voice-based tone gives information about an impending M&A event.

Table 1. Information in voice tone in the earnings call immediately preceding an M&A announcement about firm performance metrics the fiscal year after an M&A announcement.

Firm Performance Metric	r	p
net profit	0.163	<<0.001
price/operating ratio	0.217	<<0.001
price/earnings ratio	0.212	<<0.001
return on asset (ROA)	0.114	<<0.001
return on equity (ROE)	0.100	<<0.001
Tobin's Q	0.212	<<0.001
cash/liability ratio	0.189	<<0.001
debt/asset ratio	0.211	<<0.001

3.4.4. Voice Tone Predicts Future Firm Performance

We then used the methods described in Section 3.2 to determine to what extent an executive's voice tone during the earnings call immediately prior to an M&A announcement holds information about future firm performance of the acquirer or target. The results are shown in Table 1. They indicate that voice tone has robust (r=0.1-0.21) and statistically significant (p<<0.001) information about all eight performance metrics that we evaluated.

3.4.5. Voice Tone Predicts Future Cumulative Abnormal Returns

We then determined to what extent an executive's voice tone during the earnings call immediately prior to an M&A announcement holds information about cumulative abnormal returns after the announcement. The results are shown in Table 2. They indicate that voice tone does not hold much information about cumulative abnormal returns in the days and weeks after an announcement, but holds robust information (up to r=0.093) after at least a quarter after the announcement. We hypothesize that the tone of the voice of the executive reflects, in part, the executive's internal assessment of the prospects of the merger. Because only limited information currently informs pricing during M&A events, and because none of that information is based on voice tone, voice tone cannot explain the relatively immediate reaction of the market. However, after several months, the executive's internal assessment during the call will manifest itself in firm performance, which in turn leads to predictable cumulative abnormal returns.

Table 2. Information in voice tone in the earnings call immediately preceding an M&A announcement about cumulative abnormal returns at varying days following the announcement.

CAR(t)	r	p
1	0.020	0.037
2	-0.002	0.844
3	0.014	0.133
4	-0.003	0.725
5	0.013	0.178
6	0.023	0.015
30	0.024	0.010
60	0.015	0.123
90	0.042	<0.001
180	0.033	0.001
360	0.093	<<0.001

3.4.6. Information in Voice Tone Provides Valuable Investment Guidance in the M&A Context

In the previous section, we described the successful prediction of abnormal returns after varying amounts of time after the M&A announcement given the tone of the voice of the executive in the earnings call preceding the M&A event.

We then asked to what extent these predictions may be valuable for trading in the M&A space. Our calculations of abnormal returns already include corrections for beta/industry trend, and include splits and dividends. Thus, they allow for relatively direct inference of potential value. While it is obvious that there are many ways to use the information provided in our predictions, we can devise a simple trading strategy that would enter a certain stock if our prediction of abnormal returns suggests a positive return, and then exit at the corresponding time. Implementing this strategy, we also calculated the difference between cumulative abnormal returns indicated by the voice tone predictions and the average of all cumulative abnormal returns, which defined the Abnormal Growth Rate (in basis points/not annualized). The results are shown in Fig. 4. They confirm the results shown in the previous section in that the tone of the voice provides the most information beyond a few months after the announcement, and suggest an expected return (of using voice tone) of close to 700 basis points for an investment with a 360-day horizon.

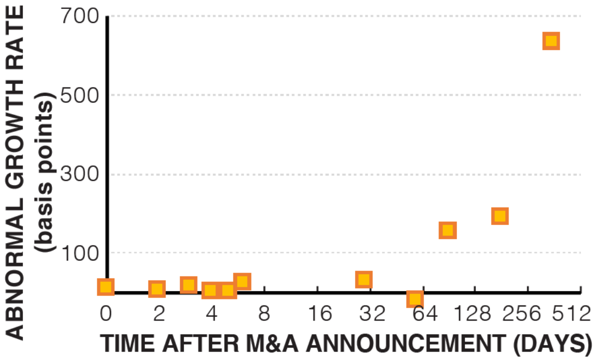

Figure 4. Voice-based tone gives valuable information.

4. MERCURY: Voice-Based Distress

4.1. Introduction

Helios Mercury provides a proprietary assessment of the distress experienced by the executive during his/her responses to the questions in the Q&A period in the earnings call. At present, Helios provides one distress measurement for each earnings call. We are also in final testing of a graphical dashboard that shows distress measurements for an executive's answer to each analyst question. As such, in contrast to COMPREHEND, which is optimized for interpretation by computers, MERCURY provides data that are optimized for interpretation by human analysts.

The distress signal provided in MERCURY is scaled between 0-5 that can be interpreted as follows:

- 0-1: low distress
- 1-2: average distress
- 2-3: high distress
- 3-4: very high distress
- 4-5: maximal distress

Progressively higher distress measurements are thought to reflect the executive's understanding that aspects of the firm are experiencing problems. Thus, it can be useful component for identifying companies that are likely going to default. Indeed, the Helios distress measurement is one of five data sources in the Default Reveal tool, a product designed by Eagle Alpha and championed by Dean Barr, the former Chief Investment Officer for Deutsche Bank. The Helios distress measurement is also useful for identifying worrysome macroeconomic trends affecting all or some industries. For example, the results shown in Section 4.2 below document that distress measurements revealed that companies are experiencing the largest levels of distress over the past decade during the ongoing COVID pandemic.

Helios MERCURY historical data are delivered as a CSV file. Its contents consist of many rows — one row for each conference call. The format of each row is described in a data dictionary provided with the CSV file.

4.2. Distress Across Time and Industries

Helios Distress characterizes the level of distress experienced by an executive during an earnings call — presumably a proxy of the level of problems facing the firm. The results shown here indicate that, when pooled across companies, the distress index identifies the level of stress experienced by firms during the ongoing COVID-19 pandemic.

Fig. 5 illustrates averaged distress values across >50k earnings calls and ˜1200 equities in North America. During the ongoing COVID-19 pandemic, firms exhibit the highest level of stress of any time over the past decade.

Fig. 6 documents the year-over-year changes (March-August 2020 vs. March-August 2019) in average distress values for all 12 industries in the Fama-French 12 Industry classification. The results suggest that industries are differentially affected by the COVID pandemic, and may also point to industries that are experiencing more general problems.

5. Conclusion

The methods, analyses, and results described in this White Paper confirm that Helios technology translates the tone of the voice of an executive during conference calls into a novel, predictive, and independent source of alpha for equities in the R1000.

6. About Helios

Founded in 2017, Helios Life Enterprises (HLE) is a pioneer in voice-based tonal analysis. HLE is the first and currently only company to conceive and devise a widely

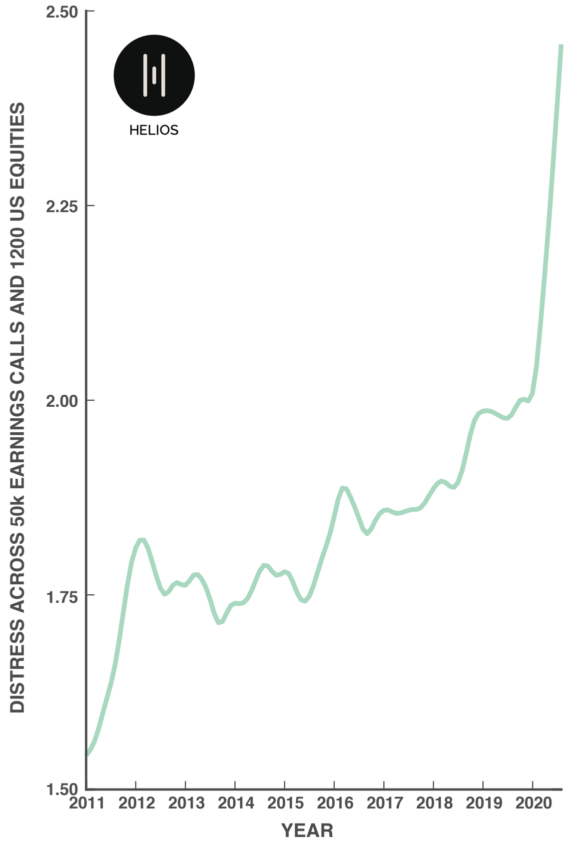

Figure 5. Average distress index across years.

available data platform that delivers systematic analytics of an executive's voice during earnings conference calls. These analytics provide novel information that is useful for predicting future earnings surprise, company performance, and cumulative abnormal returns in the context of M&A events. Core product offerings include: Helios COMPREHEND for quant researchers, and Helios MERCURY for fundamental researchers, analysts, and decision makers across all industries.

References

Alexandridis, G., Petmezas, D. and Travlos, N. G. (2010). Gains from mergers and acquisitions around the world: New evidence. *Financial Management* 39(4), 1671–1695.

Bacidore, J. M., Boquist, J. A., Milbourn, T. T. and Thakor, A. V. (1997). The search for the best financial performance measure. *Financial Analysts Journal* 53(3), 11–20.

Banse, R. and Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology* 70(3), 614.

Barrett, L. F. (1998). Discrete emotions or dimensions? The role of valence focus and arousal focus. *Cognition & Emotion* 12(4), 579–599.

Bernard, V. L. and Thomas, J. K. (1989). Post-earnings-

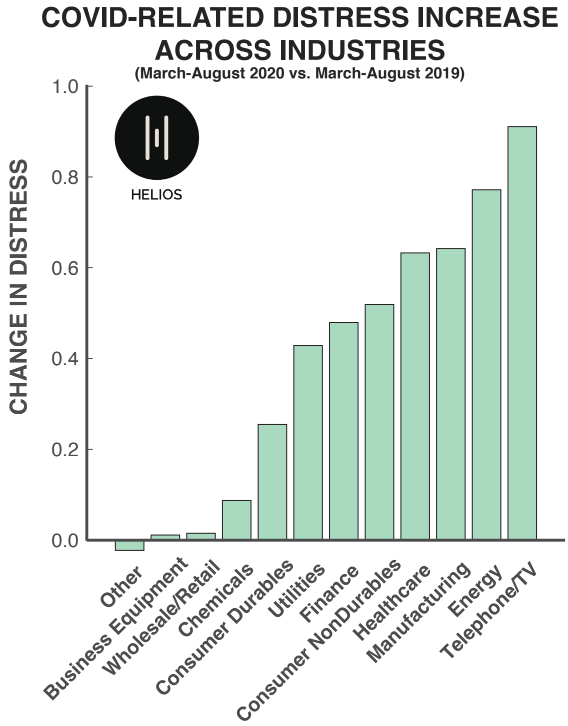

Figure 6. COVID-related distress across industries.

announcement drift: delayed price response or risk premium?. *Journal of Accounting Research* 27, 1–36.

Betton, S., Eckbo, B. E. and Thorburn, K. S. (2008). Corporate takeovers. *Handbook of empirical corporate finance* pp. 291–429.

Bharathi, S. and Geetha, A. (2017). Sentiment analysis for effective stock market prediction. *International Journal of Intelligent Engineering and Systems* 10(3), 146–154.

Bing, L., Chan, K. C. and Ou, C. (2014). Public sentiment analysis in Twitter data for prediction of a company's stock price movements. *in* '2014 IEEE 11th International Conference on e-Business Engineering'. IEEE. pp. 232–239.

Blanton, S. (1915). The voice and the emotions. *Quarterly Journal of Speech* 1(2), 154–172.

Breshears, J., Sharma, M., Anderson, N., Rashid, S. and Leuthardt, E. (2010). Electrocorticographic frequency alteration mapping of speech cortex during an awake craniotomy: case report. *Stereotactic and Functional Neurosurgery* 88(1), 11–15.

Brumberg, J. S., Krusienski, D. J., Chakrabarti, S., Gunduz, A., Brunner, P., Ritaccio, A. L. and Schalk, G. (2016). Spatio-temporal progression of cortical activity related to continuous overt and covert speech production in a reading task. *PloS One* 11(11).

Brunner, P., Ritaccio, A. L., Lynch, T. M., Emrich, J. F., Wilson, J. A., Williams, J. C., Aarnoutse, E. J., Ramsey, N. F., Leuthardt, E. C., Bischof, H. et al. (2009). A practical procedure for real-time functional mapping of eloquent cortex using electrocorticographic signals in humans. *Epilepsy & Behavior* 15(3), 278–286.

Bugental, D. B., Caporael, L. and Shennum, W. A. (1980). Experimentally produced child uncontrollability: Effects on the potency of adult communication patterns. *Child Development* pp. 520–528.

Bugental, D. B., Henker, B. and Whalen, C. K. (1976). Attributional antecedents of verbal and vocal assertiveness.. *Journal of Personality and Social Psychology* 34(3), 405.

Bugental, D. B. and Love, L. (1975). Nonassertive expression of parental approval and disapproval and its relationship to child disturbance. *Child Development* pp. 747–752.

Caffi, C. and Janney, R. W. (1994). Toward a pragmatics of emotive communication. *Journal of Pragmatics* 22(3-4), 325–373.

Campbell, J. Y. and Shiller, R. J. (1988). Stock prices, earnings, and expected dividends. *The Journal of Finance* 43(3), 661–676.

Chang, E. F., Wang, D. D., Perry, D. W., Barbaro, N. M. and Berger, M. S. (2011). Homotopic organization of essential language sites in right and bilateral cerebral hemispheric dominance. *Journal of Neurosurgery* 114(4), 893–902.

Cogan, G. B., Thesen, T., Carlson, C., Doyle, W., Devinsky, O. and Pesaran, B. (2014). Sensory–motor transformations for speech occur bilaterally. *Nature* 507(7490), 94–98.

Collins, D. W. and Kothari, S. (1989). An analysis of intertemporal and cross-sectional determinants of earnings response coefficients. *Journal of Accounting and Economics* 11(2-3), 143–181.

de Pesters, A., Taplin, A. M., Adamo, M. A., Ritaccio, A. L. and Schalk, G. (2016). Electrocorticographic mapping of expressive language function without requiring the patient to speak: A report of three cases. *Epilepsy & Behavior Case Reports* 6, 13–18.

Deng, S., Mitsubuchi, T., Shioda, K., Shimada, T. and Sakurai, A. (2011). Combining technical analysis with sentiment analysis for stock price prediction. *in* '2011 IEEE Ninth International Conference on Dependable, Autonomic and Secure Computing'. IEEE. pp. 800–807.

Drucker, H., Burges, C. J., Kaufman, L., Smola, A. J. and Vapnik, V. (1997). Support vector regression machines. *in* 'Advances in Neural Information Processing Systems'. pp. 155–161.

Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*. Chapman and Hall.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion* 6(3-4), 169–200.

Ekman, P. and Friesen, W. V. (1969). Nonverbal leakage and clues to deception. *Psychiatry* 32(1), 88–106.

Ethofer, T., Van De Ville, D., Scherer, K. and Vuilleumier, P. (2009). Decoding of emotional information in voice-sensitive cortices. *Current Biology* 19(12), 1028–1033.

Fedorenko, E., Scott, T. L., Brunner, P., Coon, W. G., Pritchett, B., Schalk, G. and Kanwisher, N. (2016). Neural correlate of the construction of sentence meaning. *Proceedings of the National Academy of Sciences* 113(41), E6256–E6262.

Fridlund, A. J. (1997). The new ethology of human facial expressions. *The psychology of facial expression* 103.

Frühholz, S., Ceravolo, L. and Grandjean, D. (2012). Specific brain networks during explicit and implicit decoding of emotional prosody. *Cerebral Cortex* 22(5), 1107–1117.

Gondhalekar, V. B., Raymond Sant, R. and Ferris, S. P. (2004). The price of corporate acquisition: determinants of cash takeover premia. *Applied Economics Letters* 11(12), 735–739.

Heaton, J. and Lucas, D. (1999). Stock prices and fundamentals. *NBER Macroeconomics Annual* 14, 213–242.

Herff, C., Heger, D., de Pesters, A., Telaar, D., Brunner, P., Schalk, G. and Schultz, T. (2015). Brain-to-text: decoding spoken phrases from phone representations in the brain. *Frontiers in Neuroscience* 9, 217.

Kim, M.-K., Kim, M., Oh, E. and Kim, S.-P. (2013). A review on the computational methods for emotional state estimation from the human EEG. *Computational and Mathematical Methods in Medicine* 2013.

Kragel, P. A. and LaBar, K. S. (2016). Decoding the nature of emotion in the brain. *Trends in Cognitive Sciences* 20(6), 444–455.

Kraus, M. W. (2017). Voice-only communication enhances empathic accuracy. *American Psychologist* 72(7), 644.

Kubanek, J., Brunner, P., Gunduz, A., Poeppel, D. and Schalk, G.

(2013). The tracking of speech envelope in the human cortex. *PloS One* 8(1).

Leuthardt, E. C., Miller, K., Anderson, N. R., Schalk, G., Dowling, J., Miller, J., Moran, D. W. and Ojemann, J. (2007). Electrocorticographic frequency alteration mapping: a clinical technique for mapping the motor cortex. *Operative Neurosurgery* 60(suppl_4), ONS–260.

Leuthardt, E., Pei, X.-M., Breshears, J., Gaona, C., Sharma, M., Freudenburg, Z., Barbour, D. and Schalk, G. (2012). Temporal evolution of gamma activity in human cortex during an overt and covert word repetition task. *Front Human Neurosci* 6, 99.

Lotte, F., Brumberg, J. S., Brunner, P., Gunduz, A., Ritaccio, A. L., Guan, C. and Schalk, G. (2015). Electrocorticographic representations of segmental features in continuous speech. *Frontiers in Human Neuroscience* 9, 97.

Loughran, T. and McDonald, B. (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance* 66(1), 35–65.

Manne, H. G. (1965). Mergers and the market for corporate control. *Journal of Political economy* 73(2), 110–120.

Martin, S., Brunner, P., Holdgraf, C., Heinze, H.-J., Crone, N. E., Rieger, J., Schalk, G., Knight, R. T. and Pasley, B. N. (2014). Decoding spectrotemporal features of overt and covert speech from the human cortex. *Frontiers in Neuroengineering* 7, 14.

Martin, S., Brunner, P., Iturrate, I., Millán, J. d. R., Schalk, G., Knight, R. T. and Pasley, B. N. (2016). Word pair classification during imagined speech using direct brain recordings. *Scientific reports* 6, 25803.

Massey Jr, F. J. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Assoc* 46(253), 68–78.

Masulis, R. W., Wang, C. and Xie, F. (2007). Corporate governance and acquirer returns. *Journal of Finance* 62(4), 1851–1889.

Mayew, W. J. and Venkatachalam, M. (2012). The power of voice: Managerial affective states and future firm performance. *The Journal of Finance* 67(1), 1–43.

Mehrabian, A. and Ferris, S. R. (1967). Inference of attitudes from nonverbal communication in two channels. *Journal of Consulting Psychology* 31(3), 248.

Mehrabian, A. and Wiener, M. (1967). Decoding of inconsistent communications. *Journal of Personality and Social Psychology* 6(1), 109.

Mehrabian, A. et al. (1971). *Silent Messages*. Vol. 8. Wadsworth Belmont, CA.

Moeller, S. B., Schlingemann, F. P. and Stulz, R. M. (2004). Firm size and the gains from acquisitions. *Journal of Financial Economics* 73(2), 201–228.

Pei, X., Barbour, D. L., Leuthardt, E. C. and Schalk, G. (2011). Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans. *Journal of Neural Engineering* 8(4), 046028.

Pei, X., Hill, J. and Schalk, G. (2012). Silent communication: toward using brain signals. *IEEE Pulse* 3(1), 43–46.

Pei, X., Leuthardt, E. C., Gaona, C. M., Brunner, P., Wolpaw, J. R. and Schalk, G. (2011). Spatiotemporal dynamics of electrocorticographic high gamma activity during overt and covert word repetition. *NeuroImage* 54(4), 2960–2972.

Penman, S. H. (1992). Return to fundamentals. *Journal of Accounting, Auditing & Finance* 7(4), 465–483.

Potes, C., Gunduz, A., Brunner, P. and Schalk, G. (2012). Dynamics of electrocorticographic (ecog) activity in human temporal and frontal cortical areas during music listening. *NeuroImage* 61(4), 841–848.

Riès, S. K., Dhillon, R. K., Clarke, A., King-Stephens, D., Laxer, K. D., Weber, P. B., Kuperman, R. A., Auguste, K. I., Brunner, P., Schalk, G. et al. (2017). Spatiotemporal dynamics of word retrieval in speech production revealed by cortical high-frequency band activity. *Proceedings of the National Academy of Sciences* 114(23), E4530–E4538.

Rodrigues, U. and Stegemoller, M. (2014). What all-cash companies tell us about IPOs and acquisitions. *Journal of Corporate Finance* 29, 111–121.

Roland, J., Brunner, P., Johnston, J., Schalk, G. and Leuthardt, E. C. (2010). Passive real-time identification of speech and motor cortex during an awake craniotomy. *Epilepsy & Behavior* 18(1-2), 123–128.

Roseman, I. J. (1984). Cognitive determinants of emotion: A structural theory. *Review of Personality & Social Psychology* .

Rosenthal, R. and DePaulo, B. M. (1979). Sex differences in eavesdropping on nonverbal cues. *Journal of Personality and Social Psychology* 37(2), 273.

Ross, E. D. (1981). The aprosodias: Functional-anatomic organization of the affective components of language in the right hemisphere. *Archives of Neurology* 38(9), 561–569.

Ross, E. D. and Mesulam, M.-M. (1979). Dominant language functions of the right hemisphere? prosody and emotional gesturing. *Archives of Neurology* 36(3), 144–148.

Routledge, B. R., Sacchetto, S. and Smith, N. A. (2017). Predicting merger targets and acquirers from text. Technical report. Working Paper, Carnegie Mellon University.

Schalk, G., Leuthardt, E. C., Brunner, P., Ojemann, J. G., Gerhardt, L. A. and Wolpaw, J. R. (2008). Real-time detection of event-related brain activity. *NeuroImage* 43(2), 245–249.

Scherer, K. R., Johnstone, T. and Klasmeyer, G. (2003). Vocal expression of emotion. *Handbook of affective sciences* pp. 433–456.

Scherer, K. R., Schorr, A. and Johnstone, T. (2001). *Appraisal processes in emotion: Theory, methods, research*. Oxford University Press.

Schwert, G. W. (2000). Hostility in takeovers: in the eyes of the beholder?. *The Journal of Finance* 55(6), 2599–2640.

Smola, A. J. and Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing* 14(3), 199–222.

Sturm, I., Blankertz, B., Potes, C., Schalk, G. and Curio, G. (2014). ECoG high gamma activity reveals distinct cortical representations of lyrics passages, harmonic and timbre-related changes in a rock song. *Frontiers in Human Neuroscience* 8, 798.

Swift, J., Coon, W., Guger, C., Brunner, P., Bunch, M., Lynch, T., Frawley, B., Ritaccio, A. and Schalk, G. (2018). Passive functional mapping of receptive language areas using electrocorticographic signals. *Clinical Neurophysiology* 129(12), 2517–2524.

Taplin, A. M., de Pesters, A., Brunner, P., Hermes, D., Dalfino, J. C., Adamo, M. A., Ritaccio, A. L. and Schalk, G. (2016). Intraoperative mapping of expressive language cortex using passive real-time electrocorticography. *Epilepsy & Behavior Case Reports* 5, 46–51.

Travlos, N. G. (1987). Corporate takeover bids, methods of payment, and bidding firms' stock returns. *The Journal of Finance* 42(4), 943–963.

Walker, M. B. and Trimboli, A. (1989). Communicating affect: The role of verbal and nonverbal content. *Journal of Language and Social Psychology* 8(3-4), 229–248.

Weitz, S. (1972). Attitude, voice, and behavior: A repressed affect model of interracial interaction. *Journal of Personality and Social Psychology* 24(1), 14.

Zuckerman, M., Amidon, M. D., Bishop, S. E. and Pomerantz, S. D. (1982). Face and tone of voice in the communication of deception. *Journal of Personality and Social Psychology* 43(2), 347.

Zuckerman, M., Larrance, D. T., Spiegel, N. H. and Klorman, R. (1981). Controlling nonverbal displays: Facial expressions and tone of voice. *Journal of Experimental Social Psychology* 17(5), 506–524.