The Tone of Voice Provides a Novel Source of Alpha

Gerwin Schalk[1]

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

Abstract

Markets are influenced in important ways by earnings conference calls. For decades, 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 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, alpha, theta oscillations

"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.

Consistent with this expectation, prior research has shown that voice tone contains information that goes beyond that contained in textual information such as sentiment, and that this information is predictive of specific equity variables (Mayew and Venkatachalam, 2012).

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 unlocks an important and previously not readily available source of alpha.

1.5. Availability of Voice-Based Alpha

Helios COMPREHEND is the first and currently only widely available data product that delivers 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

COMPREHEND 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. They are more completely and formally explicated below.

- **Feature 1 (F0 formant frequency):**
  F0 is the frequency generated by the vocal folds.
- **Feature 2 (F1 formant frequency):**
  F1 is the frequency that is modulated by the shape of the trachea.
- **Feature 3 (F2 formant frequency):**
  F2 is the frequency that is modulated by the shape of the oral cavity.
- **Feature 4 (mean acoustic amplitude):**
  The loudness of the voice.
- **Feature 5 (height of the intonation contour):**
  Fx height, according to the IPO system model (Hart et al., 2006).
- **Feature 6 (slope of the intonation contour):**
  Fx slope, according to the IPO system model.
- **Feature 7 (voice quality measure I):**
  Perceptual Speech Quality Measure according to ITU-T P.861.

- **Feature 8 (voice quality measure II):** Speech Transmission Index according to (Houtgast and Steeneken, 1971).
- **Feature 9 (minimum syllable duration):** The minimum syllable duration, an aspect of speech fluency.
- **Feature 10 (maximum syllable duration):** The maximum syllable duration, an aspect of speech fluency.
- **Feature 11 (relative number of pauses):** The number of pauses, an aspect of speech fluency.
- **Feature 12 (beat strength):** Beat strength, an aspect of speech rhythm (Tzanetakis et al., 2002).
- **Feature 13 (beat frequency):** Beat frequency, an aspect of speech rhythm (Scheirer, 1998).
- **Feature 14 (structural distortion):** Structural distortion, an aspect of pronunciation (Minematsu, 2004), measured using the Bhattacharyya Distance (Bhattacharyya, 1946).
- **Feature 15 (positional distortion):** Positional distortion, an aspect of pronunciation.
- **Feature 16 (glottis feature I):** The glottis is the part of the larynx consisting of the vocal cords and the opening between them, and affects voice modulation through expansion or contraction. Glottis feature I is the relaxation coefficient (Rd), one parameter estimated from the Liljencrants-Fant glottal model (Fant et al., 1985), estimated using the MSPD2 method (Degottex et al., 2010).
- **Feature 17 (glottis feature II):** Glottis feature II is the Function of Phase-Distortion (Degottex et al., 2011).
- **Feature 18 (neurolinguistic feature I):** Theta oscillations have been highlighted by the neurolinguistics literature (Symons et al., 2016) to represent specific affective aspects of speech, and are likely related to the rhythm of speech. Neurolinguistic feature I is the power of oscillations in the 4-8 Hz theta range.
- **Feature 19 (neurolinguistic feature II):** Neurolinguistic feature II is the phase of oscillations in the 4-8 Hz theta range.
- **Feature 20 (neurolinguistic feature III):** Neurolinguistic feature III is the instantaneous amplitude of oscillations in the 4-8 Hz theta range (Schalk, 2015, Coon et al., 2016, Schalk et al., 2017).
- **Feature 21 (syllable alternation feature):** The Modulation Transfer Function (MTF) reflects syllable alternation rate in connected speech (Houtgast and Steeneken, 1973).
- **Feature 22-26 (acoustic features):** Five acoustic features that represent the spectral behavior of the cochlea, extracted using a Gammachirp auditory filterbank (Irino and Patterson, 1997) with equivalent rectangular bandwidth (Patterson et al., 1992).

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. We reference these 26 features within and across earnings calls to account for differences in tonal characteristics across different executives. Finally, we produce three different statistical moments from the many measurements of those features during each call, resulting in 78 feature values (26 x 3) for each call.

We make these data available in two tiers: *Tier I* gives the 78 pre-processed auditory features, whereas *Tier II* gives quantitative predictions of different equity variables. The following sections give descriptions of the historical data available in these two tiers.

2.3. Historical dataset

For both *Tier I* and *Tier II*, our historical dataset covers 28433 audio recordings of earnings calls sourced from an enterprise-level data provider. They cover a period of April 2011-May 2020. The earnings calls are from a total of 1159 equities that represent the rebalanced Russell 1000 from 2011 to the present. Thus, there is no survivorship bias in the data. Also, we did not exclude any samples from our dataset for any reason. Thus, there is no selection bias in the data.

2.4. *Tier I:* Pre-processed auditory features

Tier I output of COMPREHEND provides the 78 pre-processed auditory features described above that to-

gether describe the emotional state of the executive during an earnings call. These features are predictive of different equity variables such as those described in the following section. Because *Tier I* data require further modeling/analyses to translate them into specific predictions, they are most useful to quantitative researchers.

In contrast to forward-looking data, which are delivered through our API, *Tier I* 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.

2.5. *Tier II:* Predictions of equity variables

2.5.1. Introduction

Tier I output of COMPREHEND provides preprocessed voice features. *Tier II* data, which contain predictions of specific equity variables, are derived from *Tier I* data only. Thus, they are solely derived from the tone of the voice of executives during earnings calls rather than from well known equity variables, such as book-to-market ratio, that are readily available to others. (Inclusion of those known equity variables substantially further improves predictions described below.)

2.5.2. Model Generation

The output of the models described herein are predictions of:

1. Immediate market volatility. Defined as the absolute value of the change (in %) in stock price between the close of market the day after the earnings call and the close of market the day of the earnings call.
2. Earnings surprise 1 for following quarter. Defined as (actual return-mean analyst forecast)/standard deviation of analyst forecast.
3. Earnings surprise 2 for following quarter. Defined as (actual return-median analyst forecast)/standard deviation of analyst forecast.
4. Percent uprevision for following quarter. Defined as the fraction (in %) of analysts that increase their estimates.

We used support vector regression (Drucker et al., 1997, Smola and Schölkopf, 2004) to establish the relationship between an executive's audio features during the conference call (i.e., *Tier I* data) with each of the four equity variables described above, 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, thereby forming the content of our *Tier II* data.

2.5.3. Model Evaluation

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

2.5.4. 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.

As with *Tier I* data, *Tier II* 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.

For some earnings calls, the actual target variables were not available. In this case, they are listed as 'NaN.'

[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.

The comparison of the actual and predicted equity variables in the historical *Tier II* suggests that our predictions of the four equity variables is robust (immediate volatility: $r = 0.057, p << 0.001$; earnings surprise 1: $r = 0.050, p << 0.001$, earnings surprise 2: $r = 0.052, p << 0.001$; percent uprevision: $r = 0.011, p < 0.1$.

Most importantly, our research suggests that our data are clearly orthogonal to other well-known variables, because our predictions survive even the most stringent control analyses (controlling for several equity variables and for quarter/industry fixed effects).

3. Conclusion

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

References

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.

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

Bhattacharyya, A. (1946). On a measure of divergence between two multinomial populations. *Sankhyā: the Indian Journal of Statistics* pp. 401–406.

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.

Coon, W., Gunduz, A., Brunner, P., Ritaccio, A., Pesaran, B. and Schalk, G. (2016). Oscillatory phase modulates the timing of neuronal activations and resulting behavior. *NeuroImage* 133, 294–301.

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.

Degottex, G., Roebel, A. and Rodet, X. (2010). Phase minimization for glottal model estimation. *IEEE Transactions on Audio, Speech, and Language Processing* 19(5), 1080–1090.

Degottex, G., Roebel, A. and Rodet, X. (2011). Function of phase-distortion for glottal model estimation. *in* '2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)'. IEEE. pp. 4608–4611.

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.

Fant, G., Liljencrants, J. and Lin, Q.-g. (1985). A four-parameter model of glottal flow. *STL-QPSR* 4(1985), 1–13.

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.

Hart, J., Collier, R. and Cohen, A. (2006). *A perceptual study of intonation: an experimental-phonetic approach to speech melody*. Cambridge University Press.

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.

Houtgast, T. and Steeneken, H. J. (1971). Evaluation of speech transmission channels by using artificial signals. *Acta Acustica united with Acustica* 25(6), 355–367.

Houtgast, T. and Steeneken, H. J. (1973). The modulation transfer function in room acoustics as a predictor of speech intelligibility. *Acta Acustica United With Acustica* 28(1), 66–73.

Irino, T. and Patterson, R. D. (1997). A time-domain, level-dependent auditory filter: The gammachirp. *The Journal of the Acoustical Society of America* 101(1), 412–419.

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.

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.

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.

Minematsu, N. (2004). Yet another acoustic representation of speech sounds. *in* '2004 IEEE International Conference on Acoustics, Speech, and Signal Processing'. Vol. 1. IEEE. pp. I–585.

Patterson, R., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C. and Allerhand, M. (1992). Complex sounds and auditory images. *in* 'Auditory physiology and perception'. Elsevier. pp. 429–446.

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.

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.

Schalk, G. (2015). A general framework for dynamic cortical function: the function-through-biased-oscillations (FBO) hypothesis. *Frontiers in Human Neuroscience* 9, 352.

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.

Schalk, G., Marple, J., Knight, R. T. and Coon, W. G. (2017). Instantaneous voltage as an alternative to power-and phase-based interpretation of oscillatory brain activity. *NeuroImage* 157, 545–554.

Scheirer, E. D. (1998). Tempo and beat analysis of acoustic musical signals. *The Journal of the Acoustical Society of America* 103(1), 588–601.

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.

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.

Symons, A. E., El-Deredy, W., Schwartze, M. and Kotz, S. A. (2016). The functional role of neural oscillations in non-verbal emotional communication. *Frontiers in Human Neuroscience* 10, 239.

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.

Tzanetakis, G., Essl, G. and Cook, P. (2002). Human perception and computer extraction of musical beat strength. *in* 'Proc. DAFx'. Vol. 2.

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.