

A Reply to
Money Talks: The Power of Voice. A critical review of Mayew and Venkatachalam's
The Power of Voice: Managerial Affective States and Future Firm Performance

William J. Mayew
Mohan Venkatachalam

Duke University – Fuqua School of Business

Introduction

This paper is a response to Lacerda's (2012) critique of Mayew and Venkatachalam (2012). Lacerda (2012) asserts that the results in Mayew and Venkatachalam (2012, hereafter MV) cannot be relied upon because the paper uses measures derived from Layered Voice Analysis (LVA) software. This assertion is based on Lacerda's view that any evidence based on LVA measures is untenable because LVA itself is methodologically flawed. In this paper, we argue that Lacerda's conclusion is quite premature. However, we do agree with Lacerda that more research is needed to validate measures that can capture affect either by LVA technology or any other technology.

A Reply to Lacerda (2012)

The main hypothesis tested in MV is whether positive and negative affect exhibited by managers during earnings conference calls are reflected in investors assessments of stock prices and analyst earnings estimates. Lacerda's (2012) criticism of MV stems not from his concerns about the hypothesis tested in MV *per se* as he agrees that the main objective of the paper is "a difficult but fully legitimate goal." Rather, the main concern relates to MV's use of LVA technology to derive affect measures. That is, Lacerda's claim is that the LVA technology developed by Nemesysco Inc. is "unverified (and indeed, unverifiable)." Lacerda bases this assertion exclusively on his scientific evaluation of Nemesysco's patent documents. Our reading of Lacerda (2012) suggests that the broad arguments advanced are similar to those in his earlier papers Eriksson and Lacerda (2007) and Lacerda (2009). Based on the patent document described in Liberman (2003), Lacerda (2012) asserts that the base LVA variables described in the patent document, are based on 'thorns' and 'plateaus', which are "merely counts, averages and a spread measure of meaningless events." As a result, Lacerda concludes that any variable

LVA produces cannot by definition contain meaningful or relevant information from speech.

While we do not dispute the theoretical arguments made in Lacerda (2012), we are surprised that the paper did not contain a systematic and scientific evaluation of the variables produced by LVA. The analyses in Lacerda (2012) regarding the speech waveforms and contours of a speech sample offer theoretical arguments for why LVA's variables, that are purportedly based on thorns and plateaus, can miss important attributes of speech and could potentially be influenced by the digitization process of speech. However, the theoretical arguments stem primarily from the patent document that arguably offers only a broad outline of the software's inner workings. It is hard to imagine that for-profit firms would completely reveal their trade secrets in patent applications, in the interest of protecting competitive advantages (Burk and Lemley 2002). Moreover, Nemesysco has publicly stated that the patent documents cover roughly 5% of the current technology. Therefore, in our view, using partial software code presented in a patent that is not fully representative of the software is an insufficient means of assessing whether LVA variables have any merit.

Nonetheless, concluding (from an analysis of patent documents, or on other grounds) that the LVA software does not extract relevant information from the speech signal suggests a testable hypothesis. In other words, if the positive and negative emotion variables derived from LVA used in MV are randomly generated as Lacerda (2012) implies, then we should observe no systematic relationship (or at least no replicable systematic relationship) between LVA variables and financial outcome variables. Furthermore, the LVA variables should exhibit no association with standard acoustic features of voice commonly studied in emotion research. The evidence presented in MV suggests that this is not the case. The financial outcome variables are associated with the LVA variables in the predicted directions. For example, a company's stock

return around an earnings conference call is positively (negatively) related to the positive (negative) emotions exhibited by executives as captured by the LVA variables. While one could argue that the relationship between stock prices and LVA emotion variables could obtain by statistical chance, subsequent to the publication of this paper we have conducted out of sample tests and obtain findings similar to those reported in MV. This offers reassurance that it is unlikely that the evidence presented in MV is due to random chance.

As social scientists, we are concerned ourselves about the construct validity of the variables that we use in our empirical analyses. This is precisely the reason we devoted an entire section in MV titled “Reverse Engineering the LVA Black Box” to the examination of whether the LVA metrics show systematic association with primitive acoustic variables (see Table VII in MV). If the LVA variables were mere random representations and without merit, we should observe no association with meaningful acoustic correlates of emotion. The findings in that table suggest significant systematic correlations with standard acoustic features suggesting that LVA does capture established vocal attributes. The results in Table VII of MV hold consistently in random subsets of the data as well, suggesting that the associations we document in MV are not simply a random artifact of the pooled sample we analyzed.

Despite our concerted efforts in attempting to reverse engineer the inner workings of the LVA we admit in the MV study that we cannot definitively conclude that the LVA metrics are devoid of measurement error. Therefore, we place several caveats in that study. For example, on page 12 of MV we state “We therefore caution the reader that our tests are ultimately joint tests of the hypothesis that market participants react to managerial affective states and that we capture affective states through the measures generated by the LVA software.” On page 31 of MV, we highlight the importance of additional work necessary to examine the validity of the

LVA software by stating that “We leave for future research a more thorough analysis of the relation between these and other LVA metrics with a broader set of acoustic features”.

Returning to the issue of construct validity, in Hobson et al. (2012) we conduct validity tests for one of the LVA measures from MV in a laboratory setting. In that paper, we conduct an experiment in which we generate a sample of truth-telling and misreporting subjects by providing them with incentives and opportunity to ‘cheat’ in a mock SAT test. We then induce and invoke cognitive dissonance in the misreporting subjects – the very emotional construct that one of the LVA measures purports to capture. Finally, we interview the subjects and extract LVA based voice dissonance markers and correlate them with several other established dissonance markers such as confessions of misreporting, extreme test scores, and participant’s belief revision about their own performance. We find that the LVA based dissonance marker is related to each of these markers in the predicted direction.

Furthermore, in Hobson et al. (2012) we utilize the same data as in MV and conduct tests that could not be executed at the time the analysis in MV was conducted. That is, one can view the archival evidence in Hobson et al. (2012) as further tests of one of the variables used in MV that should only hold if (i) the conclusions in MV are valid and (ii) the LVA negative emotion metric has construct validity. Hobson et al. (2012) find that the negative emotion variable is associated with financial accounting restatements, which occur on average 420 days after the time period for which the voice files are created. Moreover, this variable is only predictive of those financial restatements that result from purposeful managerial intervention, which is a necessary condition for having the negative emotion LVA is purported to capture. This, in our view, offers further construct validity for one of the measures produced by LVA. For a similar construct validity exercise of other LVA variables in a different setting, see Elkins and Burgoon

(2010).

Conclusions

We assert that Mayew and Venkatachalam (2012) is about “The Power of Voice” and not “The Power of LVA”.¹ Of course, the evidence in MV is codependent on the power of LVA. In other words, if the affect measures produced by the LVA technology are not meaningful acoustic correlates of emotion then we would not find consistent evidence in support of our hypothesis. That said, we view Lacerda’s (2012) critical evaluation of the LVA software to be exclusively from his evaluation of the information contained in the patent documents submitted by Nemesysco, the software manufacturer. While we do not completely dismiss the plausibility of his arguments, for scientific research to progress we require tangible evidence to support or negate these claims. In a related prior work, Eriksson and Lacerda (2007) appeal to early reliability studies of LVA (Hollien and Harnsberger (2006) and Damphousse et al. 2007) in order to support the claims that the LVA technology cannot by design produce meaningful outputs. Our papers, MV and Hobson et al. (2012), taken together represent another effort using empirical and experimental data to offer evidence on the validity and usefulness of LVA measures.

The view of Lacerda is that any additional research beyond that noted in Eriksson and Lacerda (2007) is simply not needed. We do not share this view, as a lack of results in scientific research can stem from numerous factors, including insufficient statistical power to observe hypothesized effects. We are studying ecologically valid speech samples of Chief Executive Officers, which we believe is a high powered setting to perform an analysis. Naturally, the inherent limitation of observing statistical associations, such as the associations we observe in MV, is that the associations could be spurious, but we view this as unlikely given the collection of empirical evidence we have observed. We are then left with the question of why we observe

¹ We do not have any financial interests in Nemesysco, Inc.

the systematic associations we observe. More research is needed to draw more definitive conclusions and draw deeper insights, and the extent to which this can occur will in the long run depend on the degree to which Nemesysco makes its algorithms available.

An additional view forwarded by Lacerda (2012) is that scientific research should not be undertaken unless the researcher knows with certainty how variables are derived. We do not share this view either, as in finance and accounting research, it is common to consider variables that purport to capture constructs of interest without knowing the precise construction of such variables. For example, Daines et al. (2010) assess how well proprietary commercial governance ratings perform. Price III et al. (2011) examine how well proprietary commercial accounting risk metrics perform. In neither study are the inner-workings of the variables known as they are by definition propriety, but that does not stop examination from happening. Our research falls more into this tradition in finance and accounting research. This is not to say we are not interested in learning why LVA works the way it does, and we hope that Lacerda and others will bring tangible evidence to bear on this issue.

References

- Burk, D. L., and M.A. Lemley. 2002. Is patent law technology specific? *Berkeley Technology Law Journal*, Vol. 17, pp. 1155-1206.
- Daines, R.M., I.D. Gow, D. F. Larcker. 2010. Rating the ratings: How good are commercial governance ratings? *Journal of Financial Economics* 98(3): 439-461.
- Damphousse, K. R., L. Pointon, D. Upchurch and K.M. Rebecca. 2007. Assessing the validity of voice stress analysis tools in a jail setting. Final Report to the United States Department of Justice.
- Elkins, A., and J. Burgoon. 2010. Validating vocal analysis software to assess credibility in interpersonal interaction: A multilevel factor analytic approach. Working Paper, University of Arizona.
- Eriksson, A., and F. Lacerda. 2007. Charlatanry in forensic speech science: a problem to be taken seriously. *International Journal of Speech, Language and the Law* 14(2), 169-193.
- Hollien, H., and J.D. Harnsberger. 2006. Voice stress analyzer instrumentation evaluation. Final report, (Contract: FA-4814-04-0011, CIFA, Gainesville, FL).
- Hobson, J., W.J. Mayew, M. Venkatachalam 2012. Analyzing speech to detect financial misreporting. *Journal of Accounting Research*, forthcoming.
- Lacerda, F. 2009. LVA technology - The illusion of lie detection" In P. Branderud & H. Traunmuller (Eds.), *FONETIK 2009*. Stockholm: Department of Linguistics, Stockholm University, 220-225.
- Lacerda, F. 2012. Money talks: The power of voice – A critical review of Mayew and Venkatachalam's The power of voice: Managerial affective states and future firm performance. <http://su.diva-portal.org/smash/record.jsf?pid=diva2:509721>.
- Mayew, W. J., and M. Venkatachalam. 2012. The power of voice: Managerial affective states and future firm performance. *The Journal of Finance* 67(1): 1-43.
- Price III, R.A., N.Y. Sharp, D.A. Wood. 2011. Detecting and predicting accounting irregularities: A comparison of commercial and academic risk measures. *Accounting Horizons* 25(4): 755-780.