Department of Linguistics

Stockholm University

# PERILUS 2012

## Money Talks: The Power of Voice

A critical review of Mayew and Ventachalam's
The Power of Voice:
Managerial Affective States and Future Firm Performance

Francisco Lacerda

March 2012

# Money Talks: The Power of Voice

## A critical review of Mayew and Ventachalam’s The Power of Voice: Managerial Affective States and Future Firm Performance

*Francisco Lacerda*
*Department of Linguistics*
*Stockholm University*
*SE-106 91 Stockholm*
*Sweden*
*francisco.lacerda@ling.su.se*

## Abstract

This paper argues that the results presented by Mayew and Venkatachalam in “The Power of Voice: Managerial Affective States and Future Firm Performance” (Mayew & Venkatachalam, 2011) are void because the authors used an irrelevant technology (LVA technology, Nemesysco) that cannot measure emotions conveyed by voice and therefore does not produce valid results. The paper explains why LVA technology cannot work for this problem and suggests that subscribing to this methodology damages the reputation of finance and banking research. The authors are encouraged to reanalyze their speech materials using appropriate scientific methods to measure meaningful acoustic correlates of emotions conveyed by speakers’ voices.

## Science and Fiction

When Jules Verne published his romance novel "De la Terre à la Lune" in 1865, he depicted a world that could not be realized at that time (Verne, 1866). Yet, in contrast to other tales of science fiction in which problems are solved by magic spells and extraordinary powers, Verne's romance used knowledge to address physical and technological issues rather than to perform miracles or to propose ad hoc magical solutions. Simply put, Verne presented a physically plausible dream grounded in knowledge, and his narrative triggered the fascination of an interested public whose questions about how and why things happened in that dream led to reasonable, sensible and non-circular answers. In the world of illusion, on the other hand, such questions get no other answer except for ad hoc postulates that must be accepted dogmatically and only make sense within the imagined world's own brand of logic.

From a literary perspective, magic is fascinating and entertaining. Poetry and fiction do not have to obey known physical laws, and their ability to exceed the limitations imposed by realistic constraints is often a great resource. From the perspective of science or economics, however, claims and processes must be demonstrable by public, coherent and grounded accounts. In a scientific context, questions about how things work must be asked and the answers must be documented and verifiable. Of course, science is not always correct, but in contrast to what happens in imaginary worlds, science is demanding and self-critical. Rather than the magic, private and arbitrary accounts of fantasy worlds, scientific explanations are based on general and transparent logical principles that must stand the challenges and skepticism raised by an educated public (Singh & Ernst, 2008).

## Voice analysis in The Journal of Finance

In principle, there is no limit to the scope or the type of questions that scientific research can attempt to answer. Any clearly formulated hypothesis that can be related to a meaningful measure of success can be objectively tested. In experimental research, the scientific quality of a study comes from the very transparency of the methodology and from the fact that the conclusions can be falsifiable in case the explanatory model is not true (Popper, 1992). Overcoming the challenge of falsification does not strictly prove that the conclusions are true, but the models emerge stronger as an outcome of the process. Yet, although academics are trained to formulate clear hypotheses and to identify logical flaws in arguments, scientists may nevertheless sometimes adopt methodologically dubious solutions that compromise the validity of their results. This may be a bit of an embarrassment for scientists who adopt ungrounded "magic methods," but the scientific community's exposure of methodological flaws (which normally are discovered and then halted during the peer review process) still contributes to a healthy, self-purging process from which science and honest scientists always emerge stronger.

A recent example of a "magic solution" sneaking into the academic world is provided by William J. Mayew and Mohan Venkatachalam's article in The Journal of Finance entitled "The Power of Voice: Managerial Affective States and Future Firm Performance" (Mayew & Venkatachalam, 2011). As stated in the opening sentence of their abstract, the authors intend to "measure managerial affective states during earnings conference calls by analyzing conference audio files using vocal emotion analysis software," a difficult but fully legitimate goal, provided there is indeed a reliable way of measuring emotion from the conference audio files used in the study. However, inferring a speaker's emotional state from an audio recording of her voice has proven to be a difficult task (Juslin & Laukka, 2001; Laukka, 2005). Whereas it can be shown that different emotional states often have an impact on temporal aspects of the speech signal, with respect to the range of variation of intensity and fundamental frequency, as well as on the declination of the fundamental frequency (Cowie, 2009; Schuller, Batliner, Steidl, & Seppi, 2011; Williams & Stevens, 1972), deriving the emotional state of speakers from the analysis of such phonetic variables is quite problematic (Spreckelmeyer, Kutas, Urbach, Altenmuller, & Munte, 2008). It is necessary to establish a plausible model of how and why valid measurements of phonetic variables may provide meaningful estimates of the speaker's emotional state. Regrettably, Mayew and Venkatachalam (2011) ignored this fundamental issue in their article and adopted Nemesysco's commercial solution, in spite of its lacking both plausibility and scientific grounds for its claims. Because of this unfortunate choice, Mayew and Venkatachalam's conclusions must be seen simply as void. Mayew and Venkatachalam's analysis is critically based on irrelevant measures of the speech signal that are used in a fundamentally flawed "secret" algorithm.

Good academic research must use models that can be submitted to public scrutiny and debate. Knowing the error sources involved in data measurements and how errors propagate in a model is a fundamental research procedure. Therefore, it is very surprising that Mayew and Venkatachalam bypass the demands of academic stringency and uncritically accept a company's self-promoted analysis scheme that seriously compromises their work and may discredit their academic field. Indeed, while acknowledging that "documenting a statistical association between affect states and capital market responses relies heavily on the precision with which affect is empirically measured" (p. 9, web-version), Mayew and Venkatachalam explain (p. 12) that they "choose a commercial product in LVA instead of constructing emotion metrics from vocal acoustic features directly because it is not clear from the literature which vocal emotion measurement model would be most appropriate." This is a strange rationale. If the scientific literature is not clear about how to extract emotional information from a speech signal, why should we believe the emotional analyses performed by Nemesysco's[1] commercial "magic black-box" are adequate just because the vendors say they are? And why would Mayew and Venkatachalam stand by Nemesysco's claims without hesitation?

Admittedly, Mayew and Venkatachalam are not experts in voice, but as responsible academics, they must rely on scientific methodology. Their uncritical acceptance of Nemesysco's claims and their failure to see through the inconclusive "evidence" posted at Nemesysco's website is simply unacceptable, particularly considering their publication is in a respected academic journal. Unfortunately, neither the authors nor the journal reviewers raised the pertinent questions about the validity of the LVA technology, presumably because The Journal of Finance does not typically deal with speech or voice. Yet, the central theme of Mayew and Venkatachalam's paper is transparently "The Power of Voice," and thereby the validity of their conclusions is critically dependent on the capacity of LVA technology to perform a valid analysis of speech signals and produce reliable emotion analysis.

## 1. The principles of LVA technology

So, is there any reason to believe that Nemesysco's LVA technology might be able to extract meaningful emotion information from a speech signal and even perform better than analyses based on phonetic measures of the speech signal? The sure answer is NO! According to Mayew and Venkatachalam, "In the LVA software we use, the base layer variables (technically termed SPT, SPJ, JQ and AVJ) are raw values obtained from unique measurements of the full vocal spectrum" (p. 10). Fortunately, these variables are described in the LVA patent (Liberman, 2003), and it is therefore clear that they simply cannot contain enough information to perform a trivial fundamental frequency analysis, let alone any relevant measure of affect conveyed by the speech signal (Lacerda, 2009) 2. The variables are merely counts, averages and a spread measure of meaningless events that Nemesysco calls "thorns" and "plateaus" and are measured in the digital signal, sampled at 11.025 kHz (the number of samples/second used to create the digital version of the signal) and 8 bit/sample (Eriksson & Lacerda, 2007). For comparison, a typical CD quality recording uses 48 kHz or 96 kHz and 16 or 32 bits per sample. "Thorns" and "plateaus" are detected by analyzing three consecutive samples at a point on the digital signal. If the amplitude of the middle sample in a triplet is higher or lower than both the first and the third samples by a pre-established threshold, a "thorn" is counted; if the steps between consecutive samples are less than the pre-established threshold, the

[1] http://www.nemesysco.com/
[2] What is meant by "full vocal spectrum" is difficult to know. The jargon of the patent, as well as that on the company's website, is full of "scientific-sounding" terms that are both incorrect or nonsense.

triplet is considered to be part of a "plateau" that is prolonged as long as the plateau condition is met and the total number of samples in the current plateau does not exceed a given maximum. In this initial counting of "thorns" and "plateaus," all amplitude information is lost. Because the initial measurements are meaningless, there is no way in which relevant information can be recovered by whatever "secret processing" that Nemesysco enjoys giving the impression that it has.

## 2. Emotion expression in voice

Emotions can indeed be conveyed by the speech signal. In addition to the speaker's choice of words, the melodic and rhythmic pattern of produced utterances (prosody) also provide cues concerning the speaker's state of mind. Typically, speech produced with a flat (monotone) melody conveys the impression of a sad, depressed or unengaged speaker, whereas melodic variation is often associated with the impression of a happy, excited or engaged speaker. However, there is no causal relationship between such melodic contours and the speaker's actual state of mind, both because the interpretation of prosodic cues is strongly influenced by cultural and social conventions, and also because speakers can explore stereotypical prosodic characteristics to induce fake impressions of emotional states. In addition, prosodic information is distributed and comprehended at utterance level. Prosody is the speech equivalent to melodic patterns in music, where the melody is conveyed by the combination of timing, duration and pauses associated with a sequence of notes rather than by the isolated notes. Descriptive statistics in LVA style, like counting the number of times a note is higher or lower than the preceding and following notes (as in "thorns"), cannot provide any useful emotional information. This impossibility is a fact and a consequence of physical and signal processing considerations, while the "demonstrations" of relevance provided by Nemesysco are misinterpretations of statistical artifacts. Nemesysco's technology is essentially an unstable digital "dowsing rod," controlled by irrelevant acoustic accidents. Let us see why the whole LVA notion falls down in the face of basic signal processing and phonetic principles.

## 3. Representing and analyzing speech

Consider the following example taken from the IPA samples of speech. The utterance "The North Wind and the Sun were disputing which was the stronger, when a traveler came along wrapped in a warm cloak" was spoken by a female American English speaker and recorded under good acoustic conditions[3]. The audio signal was digitized using a sampling frequency of 22 050 kHz (the acoustic signal is sampled 22 025 times per second) and 16 bit/sample (each sample, the signal's amplitude at the sampling moment, is quantized using 65, 536 steps). With these digitalization settings, speech signals are fairly well represented. Figure 1 shows the waveform (top panel), a spectrogram (truncated at 5.5 kHz, mid panel) and the pitch contour of the utterance. Time axes are displayed under each of the panels. An orthographic transcription, aligned with the waveform, is also shown under the time axes on the top and bottom panels. The gray reference lines in the pitch panel indicate the frequencies of musical notes from C2 (at about 130 Hz) to G3 (about 480 Hz), to help interpret the melody conveyed by the pitch contour in musical terms.

[3] The sound file is downloadable from http://web.uvic.ca/ling/resources/ipa/handbook_downloads.htm
Select American-English and then Narrative1.wav

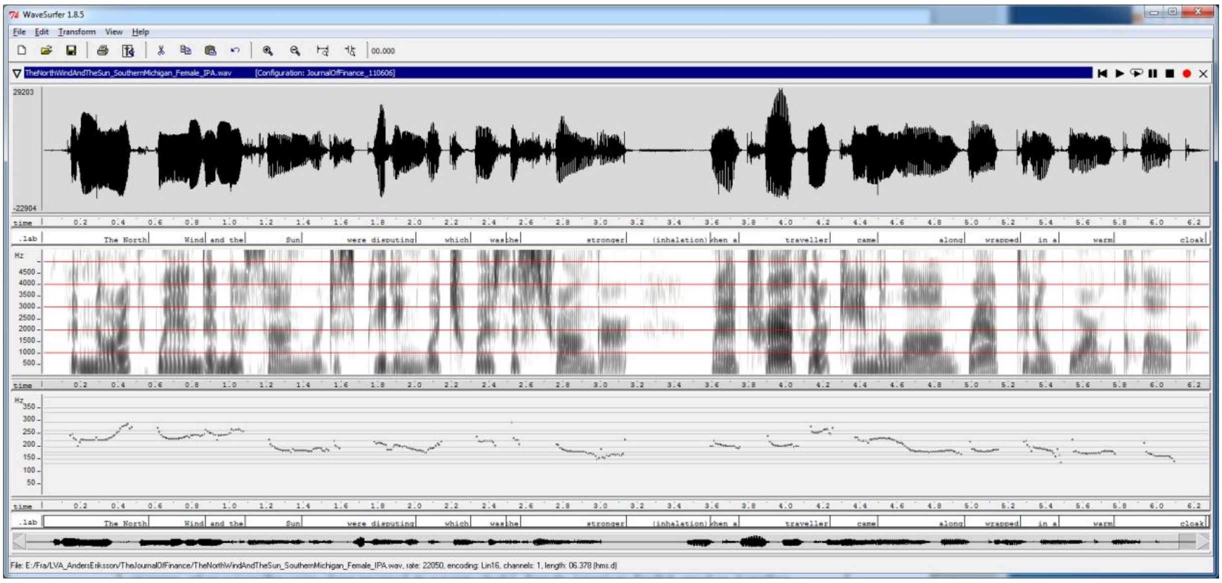


**Figure 1. Waveform, spectrogram and pitch contour of "The North Wind and the Sun were discussing which was the stronger, when a traveller came along wrapped in a warm cloak". Female American English speaker. Sampling frequency 22.050 kHz, 16 bit/sample.**

The duration of the utterance is approximately 6.2 seconds, including an inhalation pause (between 3.1 and 3.5 s), and the pitch contour has a general downward slope throughout the utterance, but it deviates from its downward trend as the speaker emphasizes some of her words. The downward pattern is also disrupted around the inhalation pause. The speaker raises her pitch slightly before the pause to indicate that she is going to continue speaking and continues the downward trend after the pause by starting at a higher frequency. This tells the listener that the last part of the utterance is a continuation of what she was saying previously. The final sharp downward movement at about 6.0 s indicates that the utterance is finished. If the pitch deviations from the downward trend throughout the utterance were less pronounced, the listener would get the impression of a sad or unengaged speaker, or conversely, for larger deviations, the speaker would be perceived as being somewhat more excited. The emotional information conveyed by the voice signal is associated with the pitch variations over rather large portions of the utterance. This is the first obvious reason why Nemesysco’s program cannot possibly detect any relevant emotional information. The program looks at 272 microseconds chunks of the signal, and these chunks are treated essentially as unrelated to each other. There is simply no way to detect any pitch movement in this span. With this length of analysis window, a single period of vocal fold vibration from a female speaker stretches over 18 such “analysis chunks.”

*A close look at the waveform*

Figure 2 displays a 600 ms segment of the original waveform, corresponding to the transition from “s” to “u” when the speaker says the word “sun.”

At this level of detail, the individual glottal pulses creating the vowel can be seen as sharp negative peaks repeating approximately every 5 millisecond (the period associated with the speaker’s 200 Hz fundamental frequency during this vowel; for a typical man’s voice, the periodicity would be about 8 ms). The glottal pulses behave like sharp percussion sounds played in a room. Their energy spreads at the speed of sound in the vocal tract walls and is reflected back and forth to create an interference pattern that is dependent on the particular geometry of the vocal tract at the moment. This extremely complex interference pattern, which can be seen between the glottal pulses in figure 2, is the vocal tract’s “sound signature” associated with perceived vowel quality. The details of pattern change inevitably as the vowel is articulated because of changes in the geometry of the vocal tract.

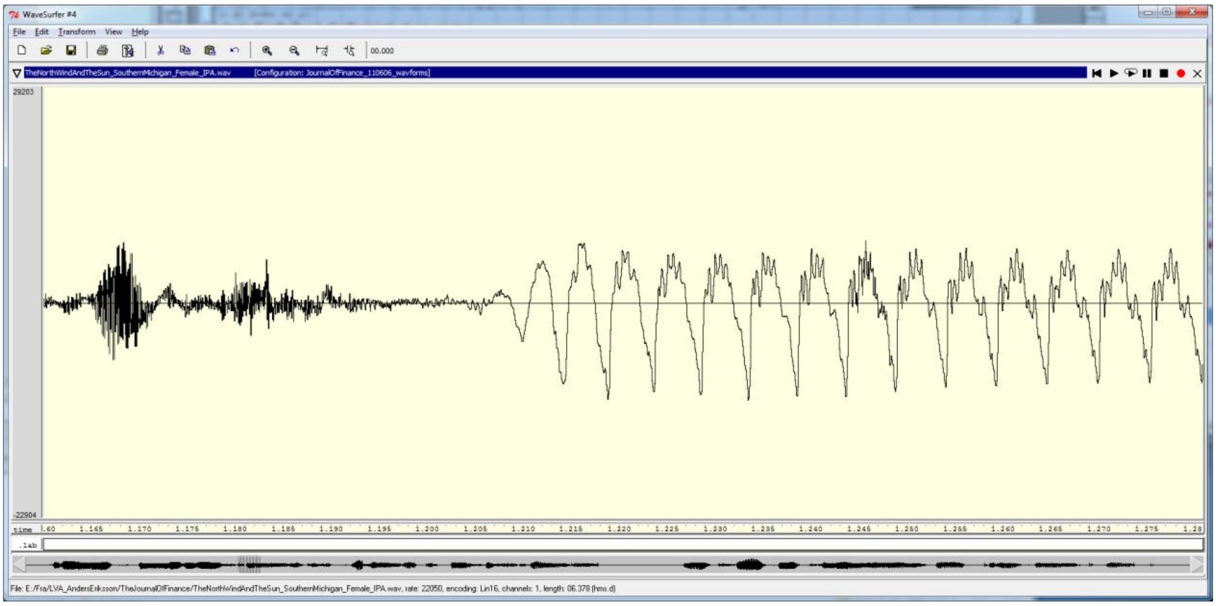

**Figure 2. A 600 ms stretch of the original waveform showing the transition from "s" to "u" in "Sun". The fricative is characterized by noise; at about 1.205 s, the vocal folds commence vibrating, and the waveform shows both their period vibration (the sharp negative peaks) and the vocal tract's response to the glottal pulses, seen in the oscillations between pulses. The pitch information is conveyed by the periodicity of the glottal pulses, while the details between glottal pulses are a consequence of how the acoustic energy of the pulses bounces back and forth as a consequence of specific articulatory configuration. Note also that the details of the waveform change along the vowel as the shape of the oral cavity changes with continuous articulatory gestures. The waveform changes because it reflects the interference between the spectral components of the sound. Different vowels will generally display different waveforms. However, when detected by a microphone, the details of a given speech sound may vary dramatically because room acoustics or external noise interferes with the original direct wave, if the signal to noise ratio is poor. Given the recording conditions and the quality of the equipment used, it is unlikely that the recordings used by Mayew and Venkatachalam would approximate the clean waveforms displayed here.**

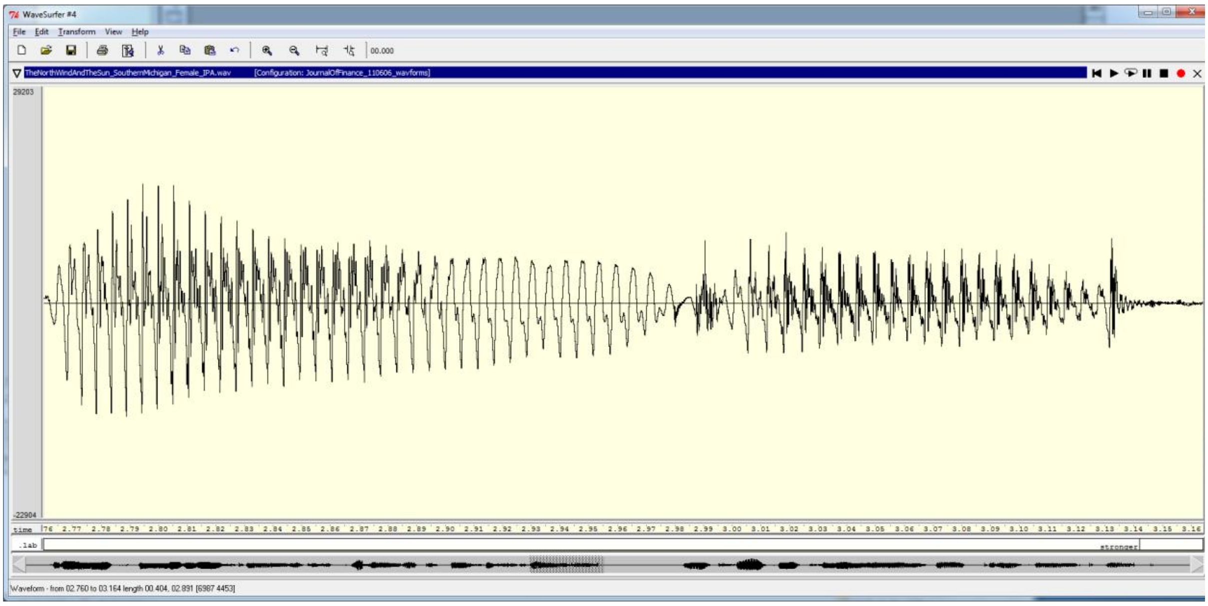

**Figure 3. Waveform from "onger" in the word "stronger". Note how the details of the waveform change in between the glottal pulses.**

Figure 3 shows another example of how the waveform changes within and across speech sounds. If the speaker had been recorded in an environment with more reverb or noise, all the spurious sounds would add to the waveforms above and further alter its details. Nemesysco's counts of "thorns" and "plateaus" are blind to these artifacts and always produce meaningless results irrespective of whether the original speech signal is clean or heavily contaminated. The output will vary in a complex way because LVA counts the amplitude accidents in the waveform, which depend both on the particular sounds being uttered and how the recording is made, but the output has no validity for the assessment of emotions. In reality, things are even worse because Nemesysco's representation of these very waveforms is so crude that it cannot even detect the details at the level at which the company would have us believe LVA is capable of performing.

## 4. The LVA "emotion analysis"

The LVA-program uses a nominal 11.025 kHz sampling rate at 8 bit/sample input, but the sample length is internally truncated to about 6.4 bit/sample after a so-called "filtering," as it is referred to in the patent. According to the program code, this "filter" is just an integer division by 3 of the original signal amplitudes. This means that the already mediocre 256 quantization levels of the original 8-bit samples are further reduced to only 85 quantization steps (the integer part of 256/3), corresponding to roughly 6 bit/sample before "thorns" and "plateaus" are measured. With this sampling frequency and sample size, the speech sounds as if it were transmitted by a bad telephone line. Amplitude changes that are smaller than a quantization step are irremediably lost. This affects both the portions of the signal that have small amplitudes (like the fricative noise "s") and the original details of the waveform in between glottal pulses. Figures 4 and 5 illustrate the direct consequences of the loss in sampling length and the (less serious) reduction of the sampling frequency in comparison to figures 2 and 3.

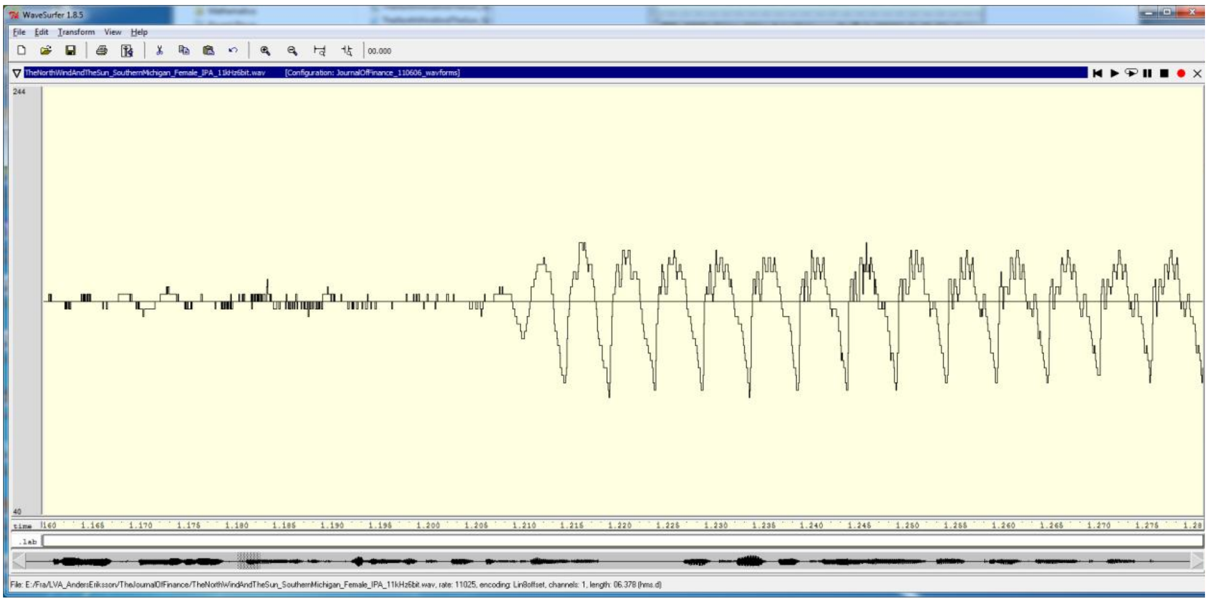

**Figure 4. A 6-bit version of the waveform displayed in figure 2.**

Interestingly, it is from this degraded representation of an audio signal that LVA's basic variables SPT, SPJ, JQ and AVJ are obtained and supposedly retrieve the minute traces of brain activity embedded in the speech signal, a notion that is, frankly, ridiculous. In the first place, there is no valid rationale for the alleged presence and expression of such traces of brain activity in the speech signal and even less for how they might be retrieved. In the second place, Nemesysco's "filter" would have damaged the alleged "minute traces" had they been present at all in the signal. "Thorns" and "plateaus" measured on this caricature of the speech wave are clearly affected by the accidents of the digitalization process. Not only are they dependent on the particular speech sounds being uttered, but they are also affected by gain, microphone characteristics, background noise and all kinds of uncontrolled acoustic disturbances, to mention only the most obvious error sources. To be sure, Nemesysco's analysis starts with a "calibration procedure" intended to establish the baseline values for further "emotion analysis." In spite of its "scientific sounding" connotations, such a calibration is just as meaningless as the rest of the analysis. Also, the calibration variables SPTcal, SPJcal, JQcal and AVJcal are dependent on the particular utterances used as well as on the accidents of the digitalization process. The differences between the calibration variables and the subsequent analyses of the rest of the recorded materials are affected by the sounds uttered under both conditions and other spurious information and have no reasonable connection to the speaker's emotional status.

In summary, a principled analysis of Nemesysco's LVA technology indicates that there are no scientific foundations for the company's claims. The technology is irrelevant and simply cannot work. It is therefore surprising how the company's aggressive marketing techniques can so successfully attract researchers who unfortunately do not ask the relevant questions and end up endorsing LVA, as in the case of Mayew and Venkatachalam (2011), who arrive at the following conclusion:

"Collectively, these results suggest that the LVA measures we use capture, at least to some extent, measurable and externally verifiable acoustic features. 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, but our exploratory analysis at a minimum rejects Lacerda's (2009) assertion that the LVA technology does not extract relevant information from the speech signal" (p. 32).

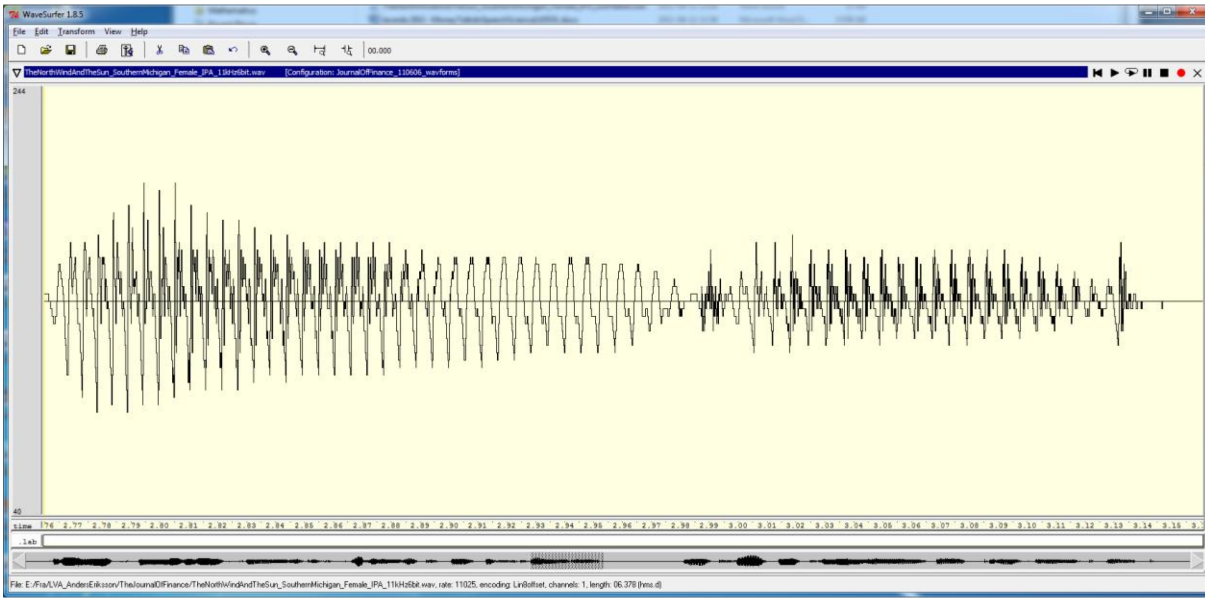


**Figure 5. A 6-bit version of the waveform displayed in figure 3.**

## 5. Unweaving the Rainbow

How can such an irrelevant technology still convince some academics and authorities that it, in fact, does something valid? The answer to this question is that the apparent strength of LVA is indeed due to the random character of the outputs it generates. On the one hand, the supposedly positive results observed are a posteriori (mis)interpretations of essentially random outputs; on the other hand, the actual level success of the LVA-based analyses is at best an extremely marginal departure from tossing a coin, which would not be accepted as positive evidence in any responsible scientific context.

*The power of random*

Indeed, because the technology performs a blind counting of many accidental sequences, it will necessarily also include some random instances where things seem to make sense, after the fact. This is a typical statistical fallacy that unfortunately even academics fail to see. An unstable and ill-defined process like LVA generates all kinds of random results that are compatible with a vast number of post hoc interpretations, including those that fit the observer's bias. For instance, because "plateaus" pick up anything that is a more or less flat section of the waveform, they will necessarily also pick up silences, as well as stop consonants and low-intensity fricative noise along with all other portions of the waveform where interference phenomena happen to have momentarily canceled out some larger oscillations. A selective interpretation of such a random output will always reveal some statistical support for vaguely defined hypotheses provided the sample of irrelevant data is large enough. There is no reasonable scientific basis for the LVA analysis, although the company has been quite successful in hiding behind a smokescreen of make-believe secret knowledge. The only reason why their products appear to work is that their algorithms function as audio-controlled random generators, with biases arbitrarily tinkered to produce "reasonable outputs." Exploring random phenomena is a standard feature of fields of pseudo-science, like the amazing "supra-natural powers" claimed by some TV entertainers (Dawkins, 1998), hiding behind the public's naïve knowledge of statistics and attraction for wonder. Isn't it strange that pseudo-scientific devices tend to "work in real life" under fuzzy assumptions of secret knowledge or special powers but fail miserably whenever strict methodological control is imposed (Damphousse, Pointon, Upchurch, & Moore, 2007; Harnsberger, Hollien, Martin, & Hollien, 2009)? Why would their performance be impaired by the "simplicity" of a controlled experiment? An answer at hand is that they never work and therefore can only convey the illusion of performance if there is enough random variation to include some outcomes compatible with some sort of vague expectation or if the performer is allowed to cheat (see e.g. "James Randi exposes Uri Geller and Peter Popoff", http://www.youtube.com/watch?v=M9w7jHYriFo).

But the problem is extremely serious when academics fail to see the fallacies of pseudo-science and even publish work based on it. Mayew and Venkatachalam (2011) should have relied on serious analysis of their speech materials. Praat, WaveSurfer or a Fast Fourier Transform may not produce meaningful results from an acoustically corrupted signal, but the sources of error are transparent, and the scientist knows the limits of the analyses. The speech signal is complex, and there are no quick fix analyses. Ignorance of the speech production process, phonetics and signal processing seems to be an illusory source of self-confidence in "solving" the problem of extracting emotions from the speech signal.

*How little "emotion" there is in the Power of Voice after all*

Finally, the meaning of the significant correlations reported by Mayew and Venkatachalam (2011) must also be questioned. Strangely, the authors accepted without question the NAFF and PAFF measures provided by Nemesysco's black box and use the correlations between these and the other variables to argue for the LVA's validity. They should know that validation by correlation is extremely unsatisfactory and that correlations definitely do not imply causality. In addition, the adjusted R-squares from their multiple regression models, supposedly endorsing the use of Nemesysco's "emotion analysis" algorithm, indicate that their models have an extremely poor explanatory value. With R-squares ranging from 5.39% to 28.8% at best, and typical values around 11-12% (average of 14.4%), most of the variance is definitely not explained by the contributions of the NAFF and PAFF variables. Indeed, this is what would be expected from an algorithm of this sort, and the observed significant correlation coefficients do not prove anything at all, unless there would be an independently motivated way of relating these arbitrarily built variables to something sensible. If a measure of the number of false alarms created by the algorithm had been used, it would be apparent that the observed correlation comes at the price of about as many false alarms as hits, since the algorithm has no basis for doing the analysis that it is claimed to perform (see e.g. Damphousse et al., 2007; Harnsberger et al., 2009).

Indeed, even if the R-squares were at some acceptable level, the fundamental issue would still be unresolved because correlations are not proofs of causality and can only be interpreted under the light of principled theoretical explanations. In other words, correlations cannot endorse Nemesysco's claims of secret, empirically-based know-how. The operational validity of Nemesysco's voice-based emotion analysis must be proven under controlled experimental conditions. Until then, it is in the public interest that academic journals live up to high review standards either by sparing their educated readership from the nonsense of the principles of LVA technology or by demanding that fundamental and relevant questions are properly answered by the inventors and users of a technology that acts as if it had some sort of scientific legitimacy.

## Summarizing and looking ahead

Responsible academic researchers must carefully check their sources and the validity of their methodological choices. If errors nevertheless occur, the scientific debate will eventually correct those using logical, transparent and grounded arguments. Regrettably, the article recently published by the finance researchers Mayew and Venkatachalam is based on Nemesysco's LVA technology, which is a voice analysis technology that has no plausible scientific basis. Indeed, the voice technology adopted by Mayew and Venkatachalam (2011) is only supported by irrelevant anecdotal or methodologically flawed studies. Such "evidence" appears to rely on the company's private "know-how," which, by definition, cannot have any scientific value. This is even reinforced by the company's public explanations of the grounds for the technology, which definitely demonstrate that it simply cannot work. Yet, Mayew and Venkatachalam accepted Nemesysco's unverified (and, indeed, unverifiable) claims rather than relying on scientifically rigorous techniques that might have been used to obtain objective and valid phonetic measures. Methodologically correct measures would have produced meaningful data that did or did not support the authors' hypotheses, but Mayew and Venkatachalam's choice of the LVA-based analysis makes all their work simply void. This might have made for a compelling and visionary approach to emotion estimation in the style of Jules Verne but is just an embarrassing fall for the appeal of a magic wonder proposed by a company that "just knows how," against all odds. Given the low esteem that the finance and banking industry currently has due to its role -- whether by design or neglect -- in the global financial crisis, subscribing to a methodology like LVA is only going to fuel more suspicion and criticism of its modus operandi.

This is a serious problem for the credibility of scientific research and may be even more damaging for the image of financial research, in particular; however, it still can be corrected by breaking with pseudo-science and reanalyzing Mayew and Venkatachalam's materials with proper speech analysis techniques. Such retesting may or may not support Mayew and Venkatachalam's hypotheses, but the results will be solid and show that phonetics raises pertinent issues that must be addressed by open knowledge from the physical, biological, psychological and information-processing domains. The sky is the limit, but phonetics embarks on a scientific expedition rather than engaging in the search for the treasure at the end of the rainbow.

## References

Cowie, R. (2009). Perceiving emotion: towards a realistic understanding of the task. [Biology]. *Philosophical Transactions of the Royal Society, 364*, 11. doi: 10.1098/rstb.2009.0139

Damphousse, K. R., Pointon, L., Upchurch, D., & Moore, R. K. (2007). Assessing the Validity of Voice Stress Analysis Tools in a Jail Setting: U.S. Department of Justice.

Dawkins, R. (1998). *Unweaving the rainbow : science, delusion, and the appetite for wonder*. Boston: Houghton Mifflin.

Eriksson, A., & Lacerda, F. (2007). Charlatanry in forensic speech science: A problem to be taken seriously. [ doi:10.1558/ijsll.2007.14.2.169 ]. *International Journal of Speech, Language and the Law, 14*(2), 169-193. doi: http://www.diva-portal.org/su/opus/publication.xml?id=13316 ; http://www.dn.se/DNet/jsp/polopoly.jsp?d=597&a=868300

Harnsberger, J. D., Hollien, H., Martin, C. A., & Hollien, K. A. (2009). Stress and deception in speech: evaluating layered voice analysis. [Evaluation Studies

Research Support, Non-U.S. Gov't]. *Journal of forensic sciences, 54*(3), 642-650. doi: 10.1111/j.1556-4029.2009.01026.x

Juslin, P. N., & Laukka, P. (2001). Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. [Research Support, Non-U.S. Gov't]. *Emotion, 1*(4), 381-412.

Lacerda, F. (2009, 2009). *LVA technology: The illusion of "lie detection".* Paper presented at the Fonetik 2009, Stockholm.

Laukka, P. (2005). Categorical perception of vocal emotion expressions. [Research Support, Non-U.S. Gov't]. *Emotion, 5*(3), 277-295. doi: 10.1037/1528-3542.5.3.277

Liberman, A. (2003). US patent Patent No. 6,638,217 B1.

Mayew, W. J., & Venkatachalam, M. (2011). The Power of Voice: Managerial Affective States and Future Firm Performance. *Journal of Finance, Forthcoming*.

Popper, K. (1992). *The logic of scientific discovery*. London: Routledge.

Schuller, B., Batliner, A., Steidl, S., & Seppi, D. (2011). Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge. *Speech Communication, 53*(9–10), 1062-1087. doi: 10.1016/j.specom.2011.01.011

Singh, S., & Ernst, E. (2008). *Trick or treatment : the undeniable facts about alternative medicine* (1st American ed.). New York: W.W. Norton.

Spreckelmeyer, K. N., Kutas, M., Urbach, T., Altenmuller, E., & Munte, T. F. (2008). Neural processing of vocal emotion and identity. *Brain Cogn*.

Verne, J. (1866). *De la terre à la lune; trajet direct en 97 heures 20 minutes*. Paris,: J. Hetzel et cie.

Williams, C. E., & Stevens, K. N. (1972). Emotions and speech: Some acoustical correlates. *Journal of the Acoustical Society of America, 52*, 1238-1250.