Voice Stress Technologies

Workshop on Field Evaluation of Behavioral and Cognitive Sciences-Based Methods and Tools for Intelligence and Counterintelligence

National Academies, Keck 100, Sep. 22-23, 2009

Philip Rubin, Ph.D.
<rubin@haskins.yale.edu>

Committee Chair and CEO, Haskins Laboratories;

Professor Adjunct, Department of Surgery, Otolaryngology
Yale University School of Medicine;
Research Affiliate, Department of Psychology, Yale University

Haskins Laboratories
THE SCIENCE OF THE SPOKEN AND WRITTEN WORD

Voice Stress Analysis (VSA)

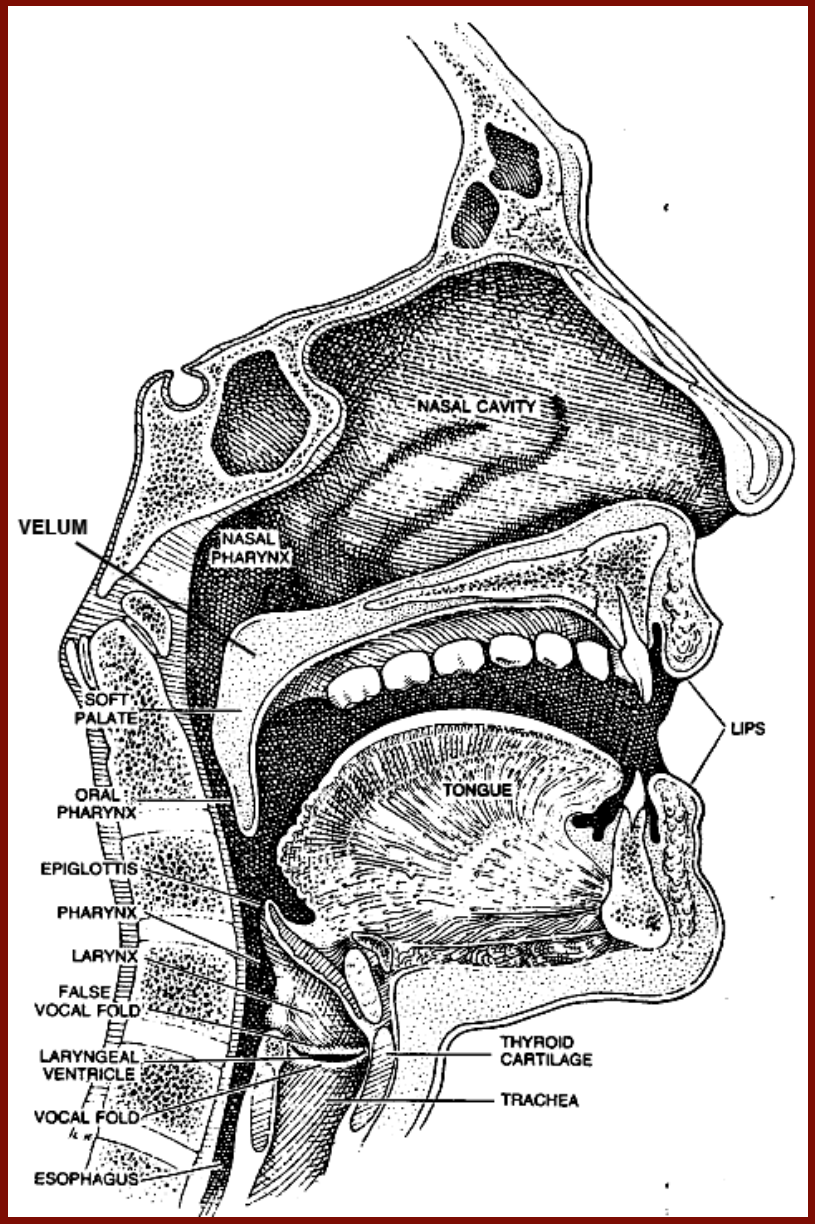
VSA techniques/technologies are based on the premise that “stress” has an inaudible, but measureable, effect on the speech signal that can be ascertained reliably and accurately using signal processing techniques.

Measurable stress is hypothesized to be an indicator of deception.

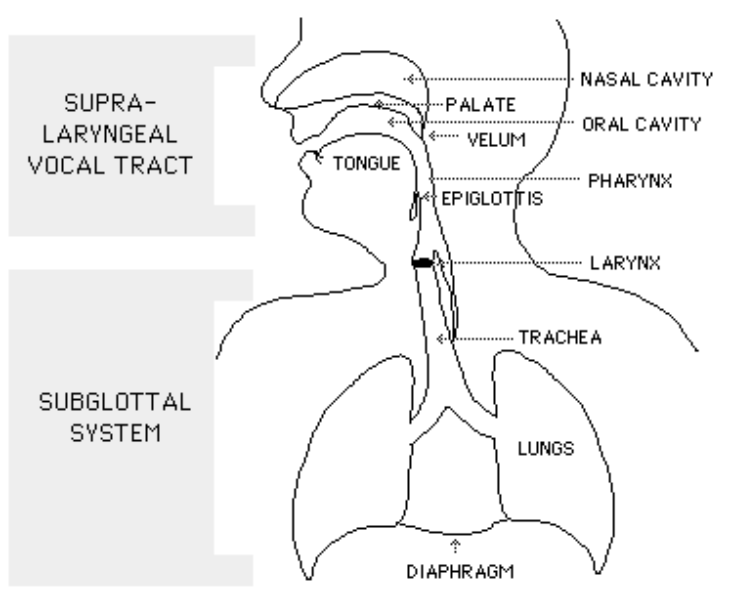
Today’s presentation will:

- Provide an introduction to how speech is produced.
- Summarize some of the main voice stress technologies.
- Discuss the testing and evaluation of voice stress technologies.
- Foster a discussion of the state-of-the-art and usefulness of VSA.

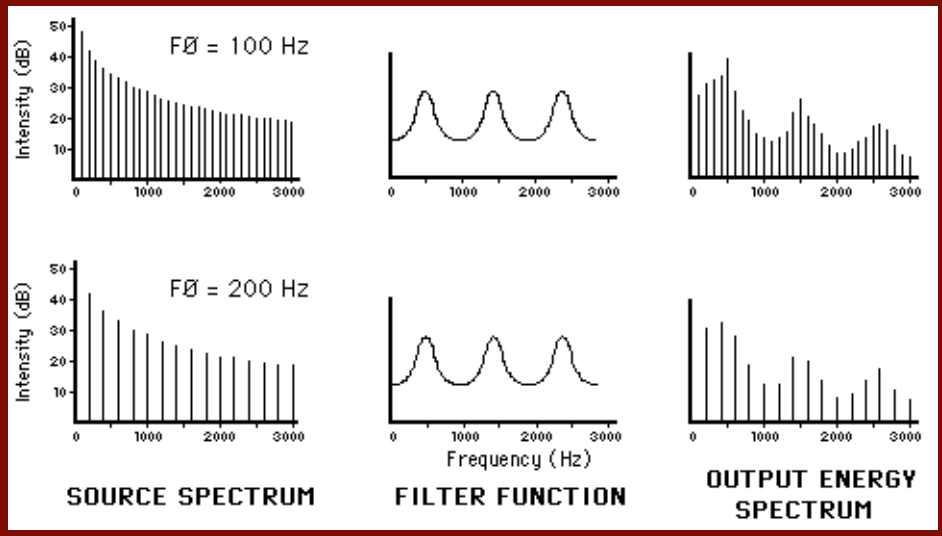
Being Biological: Speech Production



The Human Speech Production System



a.

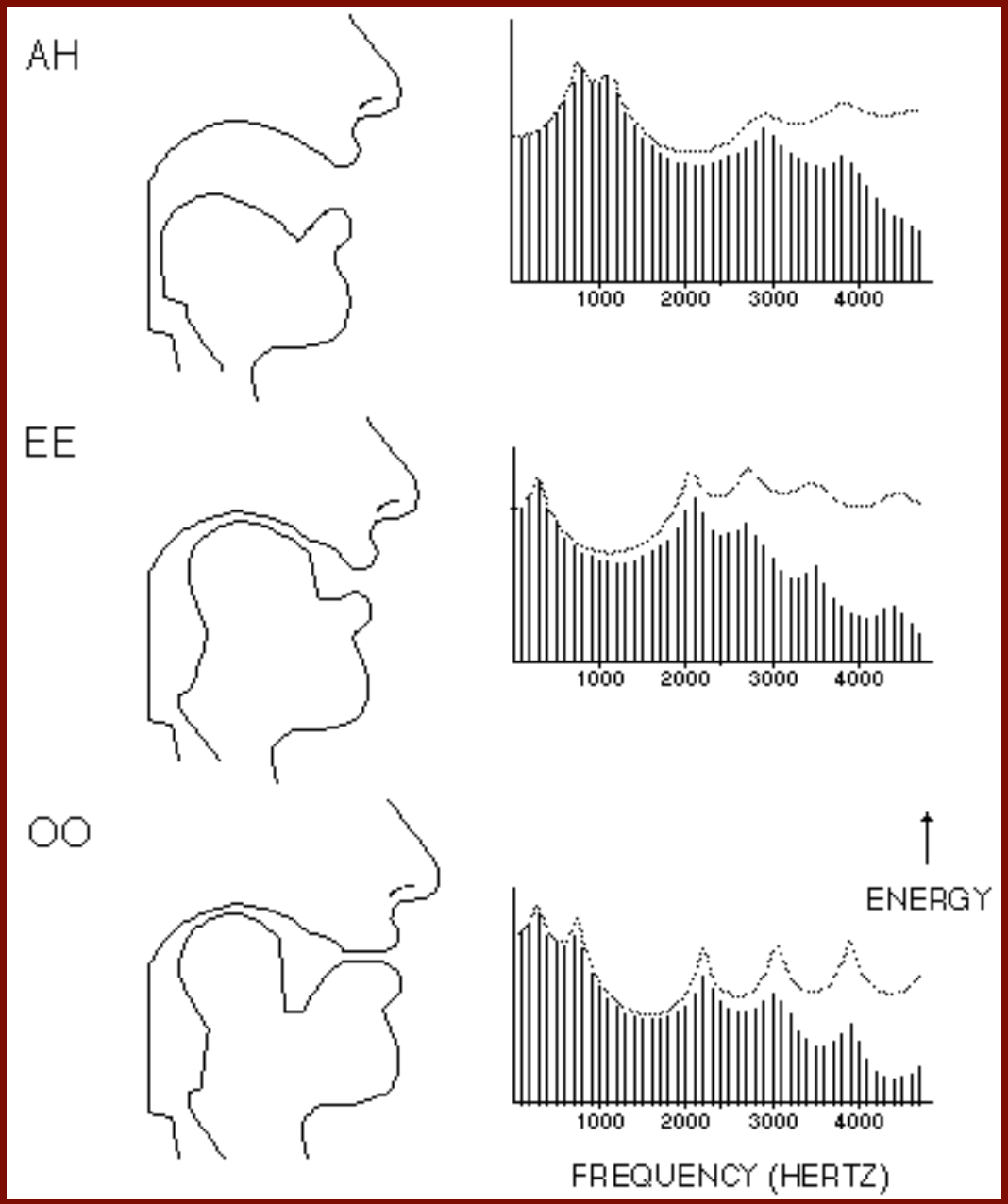


b.

a. The **supralaryngeal vocal tract**, consisting of both the oral and nasal airways, can serve as a time-varying acoustic filter that suppresses the passage of sound energy at certain frequencies while allowing its passage at other frequencies. *Formants* are those frequencies at which local energy maxima are sustained by the supralaryngeal vocal tract and are determined, in part, by the overall shape, length and volume of the vocal tract.

b. **The source-filter model of speech production**. The *source spectrum* represents the spectrum of typical glottal air flow with a fundamental frequency of 100 Hz. The *filter*, or transfer, function is for an idealized neutral vowel /uh/, with formant frequencies at approximately 500 Hz, 1500 Hz and 2500 Hz. The *output energy spectrum* shows the spectrum that would result if the filter function shown here was excited by the source spectrum shown at the left.

Source-filter model for selected vowels



Three different *vocal tract shapes* corresponding, from top to bottom, to the vowels "ah" (/a/), "ee" (/i/), and "oo" (/u/).
In this example, schematized vocal tract shapes from the Haskins Laboratories articulatory synthesizer (*ASY*) are shown.

Plotted next to each tract shape is the smoothed *transfer function* that is computationally derived by the synthesizer, and the hypothetical *energy spectrum* that would result from using these functions to filter a glottal source spectrum with a fundamental frequency of 100 Hz. Note that although all three vowels have the same fundamental frequency, their spectra differ according to the filter characteristics of the different vocal tract shapes.

Microtremor Analysis

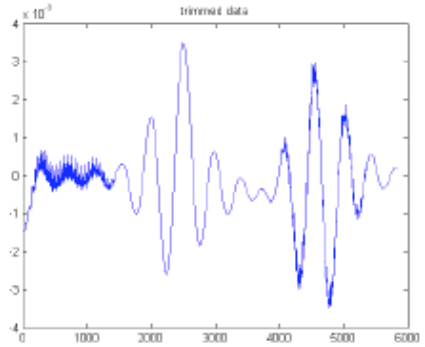


Fig.1: Neutral Sample

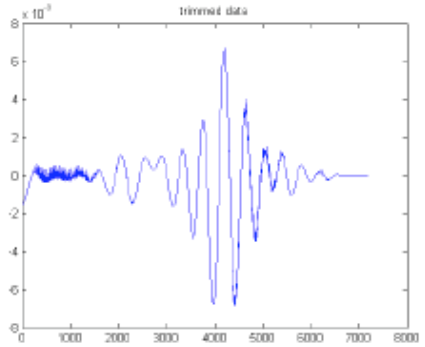


Fig. 2: High Task Stress Sample

The energy and correlation to the other five samples was then computed. Table 1 shows the energy in watts and correlation coefficient between the simulated neutral sample and the other five samples.

Table 1.

Voice Sample	Energy (Watts)	Correlation Coefficient
Neutral – Simulated	0.0079	1.0000
Low Task Stress – Simulated	0.0238	0.2278
High Task Stress – Simulated	0.0221	0.1744
Neutral – Actual	1.9046	0.1248
Low Task Stress – Actual	1.9123	0.1159
High Task Stress – Actual	1.0192	0.1636

The figures show FFT spectral analysis results. Table 1 shows the energy in watts and correlation coefficient between the simulated neutral sample and the other five samples.

From Mbitiru, N., Tay, P., Zhang, J. Z., and Adams, R. D. Analysis of stress in speech using empirical mode decomposition. *Proceedings of The 2008 IAJC-IJME International Conference,*

VSA Technologies

- Psychological Stress Evaluator (PSE)
- Computer Voice Stress Analyzer (CVSA)
- Diogenes Digital Voice Stress Analyzer – LanternPro (DDVSA™)
- Vericator™ and TrusterPro (LVA)

Review of VSA Technologies

Review of Voice Stress Analysis Based Technologies for the Detection of Deception
NOTE: TP = true positive, TN = true negative, FP = false positive, FN = false negative
April 14, 2009

Device	Conclusion	Outcomes	Year	Authors	Title	Ground Truth	Setting	Subjects
LVA	Poor Validity	TP=15% TN=95% FP=15% FN=73% Baserate=68%	2007	Damphousse, Pointon, Upchurch, & Moore OK Dept. Mental Health & Substance Abuse (NIJ)	Assessing the Validity of Voice Stress Analysis Tools in a Jail Setting	Urinalysis to determine ground truth	Prison	Inmates (n=319)
CVSA	Poor Validity	TP=8% TN=90% FP=35% FN=65% Baserate=68%	2007	Damphousse, Pointon Upchurch, & Moore OK Dept. Mental Health & Substance Abuse (NIJ)	Assessing the Validity of Voice Stress Analysis Tools in a Jail Setting	Urinalysis to determine ground truth	Prison	Inmates (n=319)
LVA	Poor Validity in detecting deception and/or stress	TP=56% TN=55% FP=45% FN=44%	2006 (part a)	Hollien & Harnsberger U. Florida (DoD)	Voice Stress Analyzer Instrumentation Evaluation	Random assignment to deception or truthful conditions	Laboratory	Adult volunteers (n=78)
CVSA	Poor Validity in detecting deception and/or stress	TP=52% TN=60% FP=40% FN=48%	2006 (part b)	Hollien & Harnsberger U. Florida (DoD)	Voice Stress Analyzer Instrumentation Evaluation	Random assignment to deception or truthful conditions	Laboratory	Adult volunteers (n=78)

From Bhatt, S. & Brandon, S., Review of voice stress based technologies for the detection of deception.

Review of VSA Technologies

Device	Conclusion	Outcomes	Year	Authors	Title	Ground Truth	Setting	Subjects
CVSA	Use of CVSA not supported by study	Below chance	2005	Cassidy	Assessing questioning protocols in detecting deception using voice stress	Mock crime where ground truth is controlled	University	College students (n = 20)
PSE	Not effective in detecting deception	Outcomes	2002	Horvath	Experimental comparison of the psychological stress evaluator and the galvanic skin response in detection of deception	Ground Truth	Setting	Subjects
Sonagram measured voice pitch, intensity & duration	Neither reliable nor useful	No better than chance	2002	Suzuki et al. Japan National Institute of Police Science	Possibility of detecting deception by voice analysis	Crime cases where truth verified via subsequent confession or medical jurisprudence	Used recorded answers to polygraph questions	Convicted criminals (n=75)
Vericator	Detected stress but not deception		2002 (part a)	Haddad, Walter, Ratley, & Smith (DOJ)	Investigation & evaluation of voice stress analysis technology	Confessions to murder	NYPD	Convicted criminals (n=2)
Diogenes	Detected stress but not deception		2002 (part a)	Haddad, Walter, Ratley, & Smith (DOJ)	Investigation & evaluation of voice stress analysis technology	Confessions to murder	NYPD	Convicted criminals (n=2)

From Bhatt, S. & Brandon, S., Review of voice stress based technologies for the detection of deception.

Review of VSA Technologies

Device	Conclusion	Outcomes	Year	Authors	Title	Ground Truth	Setting	Subjects
CVSA	Failed to detection deception	Accuracy = 38.7% (chance = 25%)	1995	Cestraro (DoD)	A comparison between decision accuracy rates obtained using the polygraph instrument and the CVSA in absence of jeopardy	Ground truth known (concealed cards)	Fort McClellan	Army recruits (n=42)
Audio pitch analysis & spectrum decomposition	No voice measure reliably indicated deception	No significant differences	1994	Cestaro & Collins (DoD)	An analysis of voice responses for the detection of deception	Ground truth known (concealed numbers)	Ft McClellan	Male military or civilian DoD employees (n=44)
Mark II Voice Analyzer	Chinese males showed higher level voice stress for prepared lies only; no effect for females	Interaction with type of lie and sex – more stress detected when discussing negative emotions about personal matters; also in males for prepared lies or personal info (not true for females)	1990	O'Hair, Cody, Wang & Chao	Vocal stress and deception detection among Chinese	Random assigned to deception of truthful conditions	University	US Chinese immigrants (n=66)
PSE	Did not distinguish concealed information	At chance levels Accuracy=30%	1980	Nachshon & Feldman	Vocal indices of psychological stress evaluator	Ground truth known (concealed cards)	University	Students (n=20)

From Bhatt, S. & Brandon, S., Review of voice stress based technologies for the detection of deception.

Review of VSA Technologies

Device	Conclusion	Outcomes	Year	Authors	Title	Ground Truth	Setting	Subjects
CVSA	Failed to detect stress	No better than chance	2000	Meyerhoff et al. (DoD)	Physiological and biochemical measures of stress compared to voice stress analysis using CVSA	HR, BP, plasma hormones, salivary samples, questionnaires (measures of stress)	Walter Reed Hospital	Army personnel (n=22)
CVSA	Failed to detect deception	Accuracy = 49.8% TP=58% TN=39% FP = 42% FN = 62% (chance = 50%)	1996	Jannine & Cestaro (DoD)	Effectiveness of detection of deception examinations using CVSA	Mock crime scenario	Ft. McClellan	Community volunteers (n=109)
CVSA	Failed to detect deception	Accuracy = 48.3% TN=71.7% TP=25.6% FN=28.9% TP=74.4% (chance = 50%)	1996	Cestaro (DoD)	A comparison of accuracy rates between detection of deception examinations using the polygraph and the CVSA in a mock crime scenario	Mock crime scenario	Ft. McClellan	Community volunteers (n=120)

From Bhatt, S. & Brandon, S., Review of voice stress based technologies for the detection of deception.

Review of VSA Technologies

Device	Conclusion	Outcomes	Year	Authors	Title	Ground Truth	Setting	Subjects
VSA	Did not distinguish deception	TP=36% (chance=33%)	1973 (part a)	Kubis (US Army Land Warfare Lab)	Comparison of voice analysis and polygraph as lie detection procedures	Random assignment to deception or truthful conditions	University	Students (n=108)
PSE	Did not distinguish deception	TP=32% (chance=33%)	1973 (part b)	Kubis (US Army Land Warfare Lab)	Comparison of voice analysis and polygraph as lie detection procedures	Random assignment to deception or truthful conditions	University	Students (n=85)
PSE	Did not distinguish concealed information	At chance levels	1973	Barland	Use of voice changes in the detection of deception	Ground truth known (concealed cards)	University	Students (n=16)

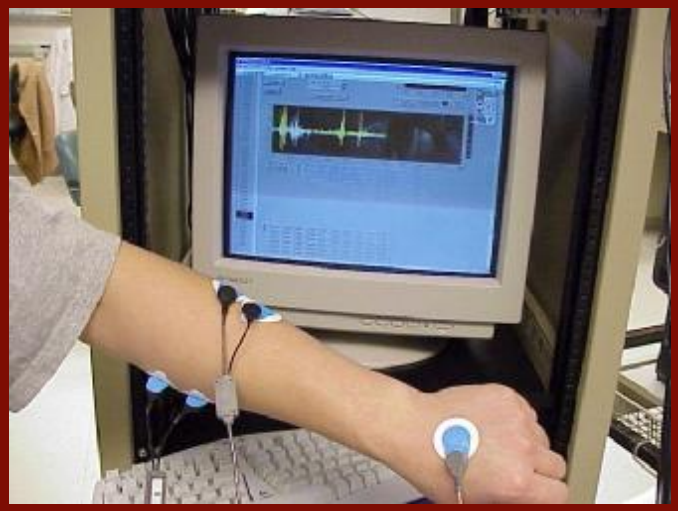
From Bhatt, S. & Brandon, S., Review of voice stress based technologies for the detection of deception.

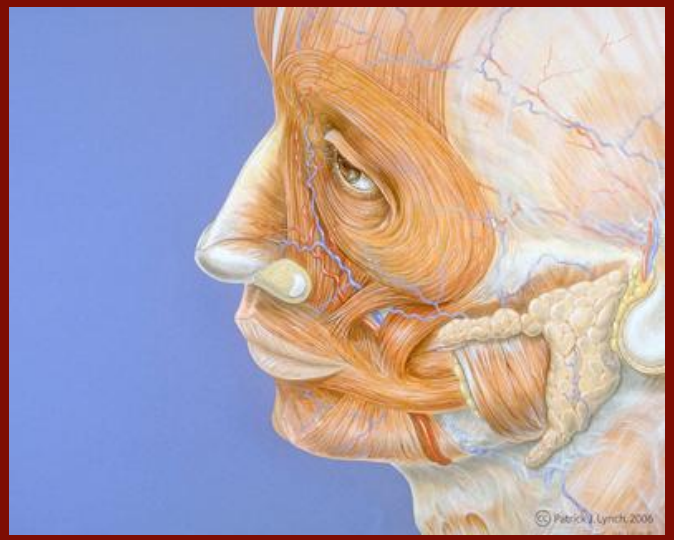
Review of VSA Technologies

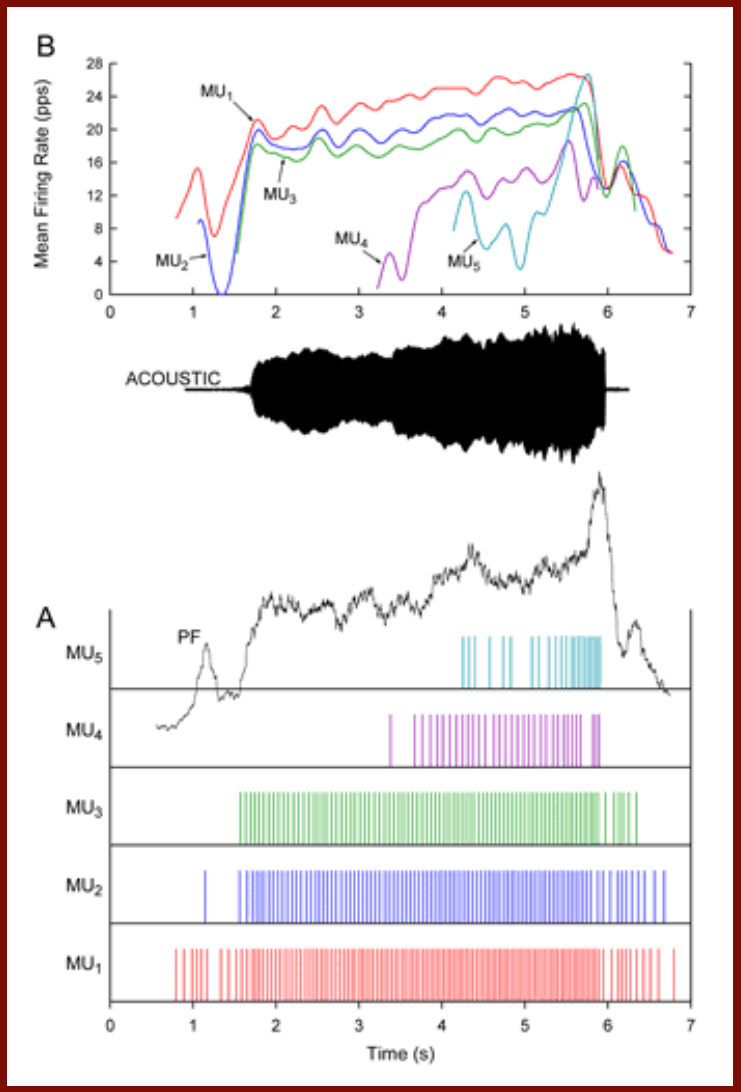
Device	Conclusion	Outcomes	Year	Authors	Title	Ground Truth	Setting	Subjects
PSE	Did not distinguish concealed information	At chance levels Accuracy=19%	1980	Nachshon & Feldman	Vocal indices of psychological stress evaluator	Criminal suspects	Police department	Criminal suspects (n=19)
PSE	Did not distinguish concealed information or stress even with increased motivation	At chance levels =20% Accuracy 18% GSR accuracy = 52%	1979	Horvath	An experimental comparison of the PSE and the GSR in detection of deception	Ground truth known (concealed cards); also measured stress with GSR	University	Students (n=64)
PSE	Did not detect deception on GKT	Below chance levels (=20%) Accuracy=15.6-21%	1979	Brenner, Branscomb, & Schwartz	Psychological stress evaluator: Two tests of a vocal measure	Ground truth known (concealed information)	University	Students (n=20)
PSE	Did not distinguish concealed information or stress	At chance levels (=20%) Accuracy=22.5% GSR accuracy =68.6%	1978	Horvath	An experimental comparison of the PSE and the GSR in detection of deception	Ground truth known (concealed cards); also measured stress with GSR	University	Students (n=60)
PSE	Did not detect deception	At chance levels Accuracy=51% (chance=50%)	1975	Barland	Detection of deception in criminal suspects: A field validation study	Criminal suspects		Criminals (n=66)

From Bhatt, S. & Brandon, S., Review of voice stress based technologies for the detection of deception.

Electromyography







Roark RM, Li JC-L, Schaefer SD, Adam A, De Luca CJ. Multiple Motor Unit Recordings of Laryngeal Muscles: The Technique of Vector Laryngeal EMG. *Laryngoscope 2002, 112:2196-2201*

Voice Stress Technologies

Issues to Consider and Discuss

- Criticality of a clear description of assumptions, methodologies, and (non-classified) algorithms and/or procedures.

- A product in hand ... The urgent need for results may overshadow the often difficult and slow process of evaluation.

- Decoupling product/scientific validity from fear/expert/geek factors (e.g. positive benefits simply from assertions of expertise and/or the “reliability” of “effectiveness” of various tools)

- The intersection of technology/product evaluation and commercial interests.

- Vehicles for quality, meaningful technology evaluation.

Haskins Laboratories

THE SCIENCE OF THE SPOKEN
AND WRITTEN WORD