REVIEW OF VOICE STRESS BASED TECHNOLOGIES FOR THE DETECTION OF DECEPTION

authors

Sujeeta Bhatt

Susan E. Brandon

December 12, 2008

REVIEW OF VOICE STRESS BASED TECHNOLOGIES FOR THE DETECTION OF DECEPTION

Executive Summary

A review of 24 empirical studies that were conducted over a period of more than 30 years failed to provide evidence of the validity or reliability of voice stress analysis-based technologies for the detection of deception in individuals. There was some evidence that some voice analysis-based technologies are able to detect voice stress reliably. A more detailed description of two recent empirical studies, both exemplary for providing independent assessments of stress and deception (Hollien, Harnsberger, Martin, and Hollien, 2008) and deception (Damphouse Pointon, Upchurch, and Moore, 2007) are described in greater detail so that some of the requisites for a good empirical study can be understood.

We had access only to website descriptions provided by voice stress analysis-based technology vendors and one field survey (conducted by DoD) for evaluation of the narratives (anecdotes) provided by field use of these technologies. While a great many of these anecdotes are likely to reflect various, commonly occurring cognitive heuristics that result in non-rational judgments, we recommend that these field reports not be ignored but that they be subjected to a more rigorous analysis.

Recommendations

1. Systematic collection of field user data by independent observers.
    a. Quantify field narrative (anecdotal) reports.
    b. Include independent assessments of stress and/or deception wherever possible.
2. Insert current voice stress analysis technologies into ongoing controlled setting studies when possible.
3. Request vendor test data prior to purchase by USG.
    a. Conduct independent assessment of test data.
    b. Construct test-bed dataset and test vendor product against this standardized test bed dataset.

Review

Voice stress analysis (VSA) refers to the use of devices that purportedly detect the existence of truth, deception, and/or stress from voice and speech. Manufacturers of VSA devices note that their low cost, short training, and high accuracy render the VSA an invaluable tool for law enforcement and military personnel. However, substantiation of VSA's utility provided by these companies has consisted of only single-case examples and/or anecdotes without any scientific evidence regarding their efficacy. Research regarding the accuracy and efficacy of VSA has, to date, been mixed at best. Although some studies have shown VSA technologies to be sensitive to stress, none of the studies found these technologies to provide reliable indicators of deception or truthfulness (for reviews, see Horvath, 2002; Krapohl, et al, 2002) Table 1 provides an overview of 24 studies conducted between 1973 and 2007. Two recent, well-designed studies that examined the efficacy of VSA, one a laboratory study (Hollien et al., 2006) and the other a field setting study (Damphouse et al., 2007) are described briefly below.

Hollien et al. (2008) evaluated the effectiveness of the Computer Voice Stress Analyzer (CVSA) in identifying stress, deception, and/or truth. A speech database consisting of voice samples, collected in the laboratory while deception and stress (via shock) levels were systematically varied, and during a field exercise at the SERE (Survival, Escape, Resistance, Evasion) school, were used as stimuli. The speech databases were created by asking the subjects to:

Sample	Description	Task
Baseline	Calibration sample	Read aloud a neutral passage several times
Sample 1	Low-stress truth	Read aloud a truthful, neutral passage
Sample 2	Low-stress lie	State low-stress false statements
Sample 3	High-stress lie	State views that strongly contradict own views under belief that friends/peers would hear – also known as harsh lies
Sample 4	High-stress truth	Read aloud a truthful, neutral passage while being shocked when the passage was read
Sample 5	Very high-stress lie	Sample 3 + 4: harsh lies + shock
Sample 6	Simulated stress	Passages read when simulating conditions of extreme stress
SERE Sample	Field data	SERE trainees were instructed to lie about training – tested using guilty knowledge test – faced punishment if deception was detected – high jeopardy

A set of University of Florida (UF) investigators trained at the National Institute for Truth Verification (NITV; manufacturers of the CVSA) and another team from NITV analyzed the randomized speech data. The NITV evaluators correctly identified 65% of the deceptive and 30% of the truthful statements (with a 70% false positive and 33% false negative rate). Similarly, the NITV team correctly identified 61% high-stress and 30% low-stress samples (with a 70% false positive and 39% false negative rate). In terms of the SERE database, the NITV group correctly identified 19% of the deceptive and 55% of the innocent trainees (with a 45% false positive and 81% false negative rate). The UF analysis team reached similar results on all three measures (deception, stress, and field data) to those of the NITV team. The authors note that when the true-positive rate is close to the false-positive rate, the device is insensitive to the measure in question (e.g, deception). The sensitivity of the device to measure deception, truth, and stress was close to zero in laboratory and field (SERE) voice databases suggesting that CVSA does not exceed chance levels in detecting deception, truth, and/or stress (Hollien et al, 2008).

In the field setting study, Damphousse et al., (2007) compared the efficiency of the CVSA and Layered Voice Analysis (LVA) devices. The investigators interviewed a random sample of 319 arrestees in an Okalahoma county jail during the booking process using CVSA and LVA to inquire about recent drug (marijuana, cocaine, opiates, PCP, and methamphetamine) use. Each interview was followed by a urinalysis to determine if the subject was deceptive regarding drug use, which allowed for comparison of VSA results to as close to "ground truth" as can be provided by urinalysis. The investigators were trained on both the LVA and CVSA devices and the data were also sent to certified CVSA and LVA examiners (experts) for their analyses. The research team compared true deception to device-indicated deception to determine validity of the VSA devices. In addition, they compared investigator to expert scoring to determine if training had an effect on accuracy. Although the LVA performed

better than the CVSA, both devices scored below chance in terms of accuracy. Only 15% of subjects who tested positive for but denied any drug use were correctly identified as deceptive by the VSA devices studied (CVSA = 8% and LVA = 21.4%). Although over 90% of the non-users were correctly identified, 10% were incorrectly classified as being deceptive. However, the use of VSA programs negatively affected deceptive responses from the subjects (e.g., the bogus pipeline effect: those who believed the VSAs to be lie detection devices were less likely to be deceptive). Further, it was determined that the research team was just as accurate as the VSA experts in arriving to their results. In sum, both CVSA and LVA show poor validity and poor sensitivity (neither efficiently identified who was deceptive), irrespective of amount of VSA training received by the examiners.

Field reports include numerous anecdotes that various voice stress analysis technologies are useful, despite the failure to show validity in the numerous studies described here. There are several possibilities to consider:

First, we might consider that these technologies at least sometimes are capable of detecting stress (although not deception), and that these are the cues that are perceived as useful in the field applications. There are some data to support the claim that CVSA and LVA discriminate stress responses (O'Hair et al., 1990). If stress occurs more often in an individual who has something to hide when he or she is confronted by law enforcement or intelligence agents than in an individual who has nothing to hide, assessment of stress may be useful. The caveat is that innocent individuals have been shown to exhibit high levels of stress when their innocence or veracity is challenged (Vrij, 2008), so that this approach should result in a high level of false positives. Further, gender, socio-cultural differences, and variations in speech sounds, rhythm/tempo, and pitch patterns of different languages can affect voice stress so caution must be taken when interpreting positive stress findings (Hollien, et al., 2008; O'Hair et al., 1990). Finally, evidence disputing CVSA's and LVA's ability to measure stress (and deception) has been identified in both laboratory and field settings and as such, positive stress findings need to be additionally examined. It is possible that the stress detected is due to the interview process (and possibly the skill of the interviewer) and not guilt-based stress (Hollein et al., 2008).

Second, there may in some instances be a *bogus pipeline effect* operating. That is, the belief that they are being tested with a "lie detection technology" makes people either avoid the test, which in itself can be an indication of some kind of guilt or desire to conceal information, or results in their confessing to instances of problematic behavior when the test occurs. This may be more likely to occur with people who are less sophisticated about technologies, and/or less familiar with psychology and what a psychology experiment might entail, so that the bogus pipeline effect should occur more often with members of the general public than with college students, and perhaps more often with members of societies with less access to technologies than Americans typically have.

These two possibilities can be assessed via systematic data collection by independent observers of the use of these technologies in the field. If voice stress analysis works because it detects stress and not deception, then independent assessments of stress (GSR, hormonal assays) should correlate with the voice stress tests. If there is a bogus pipeline effect operating, a background history of individuals tested might be useful. There may also be ways of indirectly assessing whether the individual believes in the validity of the test. It is important to emphasize that these kinds of data be collected by an independent party or agency.

There are other ways to account for the field effect anecdotes. These are to point to the pervasive and powerful decision and judgment heuristics that operate in the day-to-day

world, most often without our awareness. We are all subject to these *heuristics*, even when we try not to be. Examples of these are:

- *Confirmation bias*: Seeking evidence that supports one's favored hypothesis; recalling confirmatory evidence more easily; intentionally or unintentionally distorting evidence to make it more congruent with one's position; assigning greater weight to confirmatory sources of evidence in judgment. For example, an avid fan of CVSA remembers the times when he used this technology and got confessions. He forgets or undervalues all those times when it resulted in false positives or false negatives.

- *Hindsight bias*: This occurs when knowing a particular outcome increases the perceived probability of occurrence of that outcome. For example, a person is asked to estimate how often a VSA technology results in confessions. If he was asked this question before reading a police report on the remarkable utility of voice stress technologies ("They work 90% of the time!") and estimated 60%, when he learns that the 'actual figure' was 90%, he later recalls that his answer was 70%.

- *Imperfect memory*: This includes forgetting what happened; remembering something that didn't happen (false memories); that we 'fill in the blanks' to make sense of our experiences, and that we are easily misled or miscued when we don't remember something with confidence.

- *Availability heuristic*: When we assess the frequency or probability of an event by the ease with which instances or occurrences can be brought to mind. Since availability is a function not only of frequency or probability but also of salience, this can be problematic. For example, if confession to a brutal murder in an especially difficult and challenging case occurs using VSA technology, or if a suspect is identified as 'guilty' using VSA technology but then is released and commits an especially heinous crime, we are likely to give more weight to that single apparent success than to less spectacular (but more frequent) failures.

- *Cognitive dissonance:* When our behaviors, for whatever reason, are not aligned with our attitudes, we change our attitudes to fit our behaviors. Not the other way around. E.g., a rookie border patrol agent is told to use CVSA, which he initially distrusts. However, after 15 years experience with the device (with variable outcomes), he enthusiastically endorses this technology, despite scientific evidence that it doesn't indicate deception.

- *Technology bias*: when Americans are asked to make a judgment based on what they know or information available to them, and their judgment is not in agreement with what a computer (e.g., CVSA) says, they suspend our judgment in favor of the technology.

It is important to note that these kinds of biases occur without our awareness. However, systematic field validation of the voice stress analyses technologies would allow for an assessment of the extent to which the technologies are valid or not.

Author Notes

Sujeeta.Bhatt is a member of the faculty at Georgetown University Medical Center. Currently, she has an appointment with the Defense Intelligence Agency (DIA), DoD, serving via the Intergovernmental Position Act in the capacity of Neuroscientist in the DIA Behavioral and Social Sciences Research Program. Dr. Bhatt received her Ph.D. from American University in 2004, where her dissertation focused on the integrity of the corpus callosum in individuals following mild traumatic brain injury. In addition to receiving a postdoctoral fellowship grant from the Director of National Intelligence (2005-2007), Dr. Bhatt received a Counterintelligence Field Activity (DoD) Fellowship from the American Psychological Association in 2005. Dr. Bhatt's scientific work focuses on understanding the cognitive processes involved in deception.

SusanE.Brandon is Chief, Behavioral and Social Sciences Research Program in the Defense Counterintelligence and HUMINT Center, Defense Intelligence Agency in the U.S. Department of Defense. Prior to this, she was a senior staff member at the MITRE Corporation (2005 – 2007) and Program Chief at National Institutes of Health (NIH, 2003 – 2005). At NIH, she managed a behavioral science research program (with a focus on affect and bioregulation) for the National Institutes of Mental Health (NIMH). During her time at NIMH, she was detailed for 18 months to the Office of Science and Technology Policy as Assistant Director for Social, Behavioral and Educational Sciences. Dr. Brandon was Visiting Scientist in the Science Directorate of the American Psychological Association (2001-2003) and, for fifteen years prior to that, a faculty member in the Department of Psychology (Behavioral Neurosciences) at Yale University. She is a Fellow of the American Psychological Association and the American Psychological Society. Dr. Brandon's research is in the development and evaluation of real-time, computational models of learning and behavior.

References

Aguinis, H., Pierce, C., and Quigley, B. 1993. "Conditions under which a bogus pipeline procedure enhances the validity of self-reported cigarette smoking: a meta-analytic review." Journal of Applied Social Psychology, 23(5), 352-373.

Akers, R., Massey, J., Clarke, W., and Lauer, R. 1983. "Are self-reports of adolescent deviance valid? Biochemical measures, randomized response, and the bogus pipeline in smoking behavior." Social Forces, 62(1), 234-251.

Anolli, L., & Ciceri, R. (1997). The voice of deception: Vocal strategies of naïve and able liars. Journal of Nonverbal Behavior, 21(4), 259-284.

Barland, G. 2002. "The Use of Voice Changes in the Detection of Deception." Polygraph, 31(2), 145-153.

Board for Professional & Occupational Regulation (2003). Study of the utility and validity of voice stress analyzers. Virginia Department of Professional and Occupational Regulation.

Brenner, M., Branscomb, H., and Schwartz, G. 1979. "Psychological stress evaluator: Two tests of a vocal measure." Psychophysiology, 16(4), 351-357.

Brown, T., Senter, S., and Ryan, A. 2003. Ability of the Vericator to Detect Smugglers at a Mock Security Checkpoint. Report No. DoDPI03-R-0002. Fort McClellan, AL: Department of Defense Polygraph Institute.

Campanelli, P., Dielman, T. and Shope, J. 1987. "Validity of adolescents' self-reports of alcohol use and misuse using a bogus pipeline procedure." Adolescence, 22(85), 7-22.

Cestaro, V. 1996. "A comparison between decision accuracy rates obtained using the polygraph instrument and the Computer Voice Stress Analyzer (CVSA) in the absence of jeopardy." Polygraph, 25(2), 117-127.

Cestaro, V. 1995. A Comparison Between Decision Accuracy Rates Obtained Using the Polygraph Instrument and the Computer Voice Stress Analyzer (CVSA) in the Absence of Jeopardy. (DoDPI95-R-0002). Fort McClellan, AL: Department of Defense Polygraph Institute.

Cestaro, V. 1995. A Comparison Between Decision Accuracy Rates Obtained Using the Polygraph Instrument and the Computer Voice Stress Analyzer in a mock crime scenario. (DoDPI95-R-0004). Fort McClellan, AL: Department of Defense Polygraph Institute.

Cestaro, V. and Dollins, A. 1996. "An analysis of voice responses for the detection of deception." Polygraph, 25(1), 15-342.

Cohen, J. 1960. "A coefficient of agreement for nominal scales." Educational and Psychological Measurement, 20(1), 37-46.

Cohen, J. 1968. "Weighted kappa: Nominal scale agreement with provision for scaled disagreement of partial credit." Psychological Bulletin, 70(4), 213-220.

Crites, J. January 9, 2005. Aviation Security: Identifying Threats and Assessing Feasibility of Responses to the Aviation Group of the Transportation Research Board.

Fuller, B.F. 1984. "Reliability and validity of an interval measure of vocal stress." Psychological Medicine, 14(1), 159-166.

Gamer, M., Rill, H.-G., Vossel, G., Godert, H. 2006. Psychophysiological and vocal measures in the detection of guilty knowledge. International Journal of Psychophysiology, 60(1), 76-87.

Haddad, D., Ratley, R., Walter, S., and Smith, M. 2002. Investigation and Evaluation of Voice Stress Analysis Technology, Final Report, NCJ Number: 193832. Washington, DC: National Institute of Justice.

Hansen, J. 1996. "Analysis and compensation of speech under stress and noise for environmental robustness in speech recognition." Speech Communications, Special Issue on Speech Under Stress, 20, 151-173.

Hansen, J. and Zhou, G. 1999. Methods for Voice Stress Analysis and Classification, Final Technical Report. Rome, NY: US Air Force Research Laboratory.

Harnsberger, J. D., Hollien, H., Martin, C. A., & Hollien, K. A. (in press). Journal of Forensic Sciences.

Hollien, H., Geison, L., Hicks, J. 1987. "Voice stress analysis and lie detection." Journal of Forensic Sciences, 32(2), 405-418.

Hollien, H., & Harnsberger, J. D. (2006). Voice stress analyzer instrumentation evaluation. Final Report, Gainesville, FL: CIFA Contract FA 4814-04-0011.

Hollien, H., & Harnsberger, J. (2006). The use of voice in security evaluations. The Journal of Credibility Assessment and Witness Psychology, 7(2), 74-78.

Hopkins, C., Ratley, R., Benincasa, D., and Grieco, J. 2005. "Evaluation of Voice Stress Analysis Technology." Proceedings of the 38th Hawaii International Conference on System Sciences.

Horvath, F. 1978. "An experimental comparison of the psychological stress evaluator and the galvanic skin response in detection of deception." Journal of Applied Psychology, 63(3), 338-344.

Horvath, F. 1979. "Effect of different motivational instructions on detection of deception with the psychological stress evaluator and the galvanic skin response." Journal of Applied Psychology, 64(3), 323-330.

Horvath, F. (2002). Detection deception: The promise and the reality of voice stress analysis. Polygraph, 31(2), 96-107.

Janniro, M. and Cestaro, V. 1996. "Effectiveness of detection of deception examinations using the Computer Voice Stress Analyzer." Polygraph 27(1), 28-34.

Jones, E. and H. Sigall. 1971. "The bogus pipeline: a new paradigm for measuring affect and attitude." Psychological Bulletin, 76, 349-364.

Krapohl, D. J., Ryan, A. H., & Shull, K. W. (2002). Voice stress devices and the detection of lies. Policy Review.

Kubis, J. F. 1973. Comparison of Voice Analysis and Polygraph as Lie Detection Procedures. (Technical Report No. LWL-CR-03B70, Contract DAAD05-72-C-0217). Aberdeen Proving Ground, MD: U.S. Army Land Warfare Laboratory.

Lynch, B. and Henry, D. 1979. "A validity study of the psychological stress evaluator." Canadian Journal of Behavioral Science, 11(1), 89-94.

Meyerhoff, J., Saviolakis, G., Koening, M., Yourick, D. (2001). Physiological and biochemical measures of stress compared to voice stress analysis using the Computer Voice Stress Analyzer (CVSA). Report No. DoDPI98-R-0004. Fort Jackson, SC.

National Research Council. 2003. The Polygraph and Lie Detection. Washington, DC: National Academy of Sciences.

O'Hair, S., Cody, M. J., & Behnke, R. R. (1985). Communication Apprehension and Vocal Stress as Indices of Deception. The Western Journal of Speech Communication, 49, 286-300.

O'Hair, D., Cody, M. J., Goss, B., & Krayer, K. J. (1988). The effect of gender, deceit orientation and communicator style on macro-assessments of honesty. Communication Quarterly, 36(2), 77-93.

O'Hair, D., Cody, M. J., Wang, S., & Chao, E. Y. (1990). Vocal stress and deception detection among Chinese. Communication Quarterly, 38(2, Spring), 158 - 169.

Palmatier, J. 1999. The Computerized Voice Stress Analyzer: Modern Technological Innovation or "The Emperor's New Clothes"? General Practice, Solo & Small Firm Division Magazine (16):224-232.

Palmatier, J. 2000. The Validity and Comparative Accuracy of Voice Stress Analysis as Measured by the CVSA: A Field Study Conducted in a Psychophysiological Context. Unpublished manuscript, Michigan Department of State Police.

Palmatier, J. No date. The CVSA, Polygraph, and Trustech / Vericator Voice Analysis Technologies: Preliminary Results from a Comparative Analysis Conducted in a Criminal Justice Field Setting. Unpublished manuscript access on December 1, 2006 at http://www.nemesysco.com/vericator.pdf.

Smith, G. 1977. "Voice analysis for the measurement of anxiety." British Journal of Medical Psychology, 50, 367-373.

Sprangers, M. and J. Hoogstraten. 1987. "Response-style effects, response-shift bias and a bogus-pipeline." Psychological Reports, 61, 579-585.

Suzuki, A., Watanabe, S., Taheno, Y., Kosugi, T., & Kasuya, T. 1973. "Possibility of detecting deception by voice analysis." Reports of the National Research Institute of Police Science, 26(1, February), 62-66.

Suzuki, A., Watanabe, S., Taheno, Y., Kosugi, T., & Kasuya, T. (2002). Possibility of detecting deception by voice analysis. Polygraph, 31(2), 129-134.

Timm, H. 1983. "The efficacy of the psychological stress evaluator in detecting deception." Journal of Police Science and Administration, 11(1), 62-68.

Waln, R. and Downey, R. 1987. "Voice stress analysis: Use of telephone recordings." Journal of Business and Psychology, 1(4), 379-389.

Wylie, M. 2002. "Police Use of Voice Stress Analysis Generates Controversy." Newhouse News Service (accessed at http://polygraph.com.au/pdf/police_use_of_voice_stress_analysis_generates_controversy. pdf on December 1, 2006.

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
December 12, 2008

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=35% FP=65% FN=44%	2006 (part a)	Hollien & Harnsbeger 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 & Harnsbeger U. Florida (DoD)	Voice Stress Analyzer Instrumentation Evaluation	Random assignment to deception or truthful conditions	Laboratory	Adult volunteers (n=78)

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
Sonogram 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 b)	Haddad, Walter, Ratley, & Smith (DOJ)	Investigation & evaluation of voice stress analysis technology	Confessions to murder	NYPD	Convicted criminals (n=2)

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	Janniro & 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=71.1% TP=25.6% FN=28.9% FP=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)

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 & Dollins (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)

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=18.6-21%	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)

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)