Using Technology to Separate Truth From Lies: A Review of the History, Science, Methodology, and Legal Implications of the Computer Voice Stress Analyzer

Thomas R. Cole Criminal Justice Institute School of Law Enforcement Supervision

Ms. Samantha White

May 8, 2020

Abstract

This paper looks at the history, science, methodology, and legal implications regarding the use of technology-assisted lie detection in law enforcement, with a more in-depth look at one case that went wrong. The paper focuses on the Computer Voice Stress Analyzer Test, but includes some discussion of the Polygraph test for historical and comparative context. Ultimately, this paper argues that the CVSA is a powerful tool for law enforcement, but that like any powerful tool, its true value - or danger - is determined by the person behind the tool.

Keywords: polygraph, computer voice stress analyzer, lie detector, law enforcement

Using Technology to Separate Truth From Lies: A Review of the History, Science,

Methodology, and Legal Implications of the Computer Voice Stress Analyzert

The invention of technology-assisted lie detection can be attributed to several people from as far back as the early 1900s. William Marston began researching how deception can be detected through changes in blood pressure in 1915 as part of his graduate research in Harvard University's Psychological Laboratory (National Research Council, 2003, p. 292). John Larson and Leonarde Keeler are credited with creating machines that measured multiple physiological variables instead of only looking at blood pressure (National Research Council, 2003, p. 296). Larson also focused on developing a standardized method of constructing questions for his tests (National Research Council, 2003, p. 296).

Today, law enforcement agencies around the world utilize technology-assisted lie detection devices for both criminal investigations as well as pre-employment screening of potential applicants. The two primary technologies in use are the Polygraph Test and the Computer Voice Stress Analyzer (CVSA). Both tests are based on the idea that subjects who lie will experience stress, and that stress manifests itself as predictable, unconscious physiological changes in the body that can be measured to such a degree that they can be used to differentiate between a truthful answer and a deceptive one (NITV, 2015, pp. 37-39).

The Polygraph Test measures heart rate, breathing rate, blood pressure, and galvanic skin response, all four of which increase in a subject experiencing heightened stress levels (National Research Council, 2003, p. 32). The CVSA measures changes in the subject's voice that are produced when increased stress causes the muscles controlling the vocal cords to have a decrease in physiological tremors, which they define as "minute undulation or oscillation, which occurs in working muscle" (NITV, 2015, p. 38). According to the National Institute of Truth Verification (NITV) Federal Service's *Certified Examiner's Course Operating Manual* (2005):

Under relaxed conditions, the human voice muscles are under the control of the central

nervous system and tremor at normal frequency. The onset of stress causes the sympathetic branch of the autonomic nervous system to override the parasympathetic branch, and the muscles begin to tighten up. . . Although inaudible, a change, decrease, or absence of this phenomenon can be detected by the CVSA (p. 39).

This inaudible change in the voice is recorded by a microphone and represented visually as a chart that is evaluated by the CVSA examiner to determine if the subject is being deceptive (NITV Federal Services, 2015, p. 39).

The questions asked during a CVSA exam can not be formulated haphazardly. The need for appropriately structured questions was noted by John Larson in the early 1900s, and its importance is still recognized (National Research Council, 2003, p. 296). In the NITV operating manual, at the beginning of the chapter entitled "Question Formulation & Test Formats", they have a warning that states "If the question formulation criteria outlined in this chapter are not followed, errors in analysis of charts will [emphasis added] be made" (p. 25).

There are three types of questions asked in all CVSA examinations. These are relevant questions, irrelevant questions, and control questions. Each question has a specific purpose and specific criteria that must be met.

Relevant questions must be "direct" and "short and to the point", and should "pertain[] to the subject at hand" and "contain only one issue" (NITV, 2015, p. 25). So, for example, a question like "Did you kill Jane Doe?" meets the criteria of a good question. However, a question like "Did you break into Jane Doe's house and kill her before taking the jewelry from her jewelry box?" does not meet the criteria because it contains several issues (burglary, murder, and theft). In order to meet the criteria for well-constructed relevant questions, this

would have to be split up into three separate questions, one about each issue.

Irrelevant questions always follow relevant questions and control questions. They must be "a known truth", they must "cause no stress by [themselves]", and they must have "no connection to the issue at hand" (NITV, 2015, p. 25). They are used to detect delayed and carry-over stress. Delayed stress occurs when the subject does not show significant stress on a relevant question, but shows more stress on the question immediately following. Carry-over stress occurs when the subject shows stress on a relevant question, and that stress is strong enough to persist as the subject answers the following question. By using irrelevant questions after every relevant and control question, the test's structure mitigates the possibility that a subject will show delayed or carry-over stress on a relevant question about which they are actually being truthful (NITV, 2015). Examiners are also warned of the possibility of seemingly innocuous irrelevant questions causing stress (NITV, 2015). For example, asking a subject if they are married just because they are wearing a wedding ring may induce stress if the subject is suffering from marital difficulties. (NITV, 2015).

Control questions are irrelevant questions that the test subject is instructed to purposely lie on (NITV, 2015). All CVSA tests contain precisely two control questions, and they must each be "a known truth", "a directed lie", and have "no connection to the issue at hand" (NITV, 2015 p. 25). According to an article by the NITV on their website, control questions can be helpful by comparing the subject's response to them with the subject's response to relevant and irrelevant questions (NITV, 2017, paragraph 12). Interestingly, there is no specific guidance in the article or the NITV's *Operating Manual* on how an examiner should use the interpretation of control question results in evaluating the overall test.

When conducting a CVSA examination, significant attention is given to the question formulation phase. Examiners must be familiar enough with the issue at hand to conduct a pretest interview and determine how best to phrase the relevant questions in order to have a successful examination (NITV, 2015). In a criminal investigation, this means reviewing available case materials, including police reports, video footage, physical evidence, and any other information relevant to the case (NITV, 2015). Once familiar with the case, an examiner can conduct the pre-test interview.

In interviewing the subject prior to conducting the actual examination, the examiner has the opportunity to gain valuable insight into the subject's level of willingness to engage in the process, as well as establish norms of how the subject acts when telling known and/or verifiable truths. The examiner also uses this time to establish their own credibility and the validity of the technology and equipment. In some instances, an examiner who convinces a subject of the accuracy of the testing process may get a full confession during the pre-test interview, thus negating the necessity of completing the actual examination.

The pre-test interview is also when the relevant questions for the test are formulated (NITV, 2015). Rather than decide on questions prior to the interview, it is incumbent on the examiner to consider any new information learned during the interview in creating the questions (NITV, 2015). It is possible that a subject may admit to more than they did in previous interviews related to the case, and these additional admissions may necessitate different questions than would have been formulated prior to the interview. The formulation of questions is a collaborative process, and the test subject is fully aware of all examination questions before the exam begins (NITV, 2015). As the examiner and test subject work

together to formulate questions, the subject may give up subtle (or not so subtle) clues about the case. For example, if an examiner suggests the question "Did you show a gun to Jane Doe?" and the test subject suggests changing the question to "Did you point a gun at Jane Doe?", the examiner may want to explore the differences in those two statements and why the subject suggested the change.

Once questions are agreed upon by the examiner and the test subject, a microphone is attached to the subject's collar area and a calibration process is conducted to make sure the subject is able to answer loudly enough and that the microphone is operational. Once this is completed, the exam begins and all questions are asked. There are no surprise questions, only the questions already agreed upon by both parties. After the question and answer sequence is completed the first time, the examiner explains that the question sequence will be repeated. A second test is always performed because people telling the truth tend to have better results on a second test as situational stress decreases, whereas deceptive subjects tend to do worse on the second test as they are forced to repeat their deception, causing stress to build rather than dissipate (NITV, 2015).

After the second test is run, the examiner evaluates the chart. There are numerous types of patterns that each *yes* or *no* answer can create, and these patterns indicate varying levels of deception (NITV, 2015). There are several categories into which a response pattern can be classified - more than can be covered within the bounds of this paper. However these patterns can be more roughly classified as non-deceptive, deceptive, or having a tendency toward deception (NITV, 2015). If all relevant questions are classified as non-deceptive on the second test, the test and subject are considered to be non-deceptive (NITV, 2015). If any

relevant questions are classified as deceptive (including irrelevant questions that show delayed stress after a relevant question), the test and subject are considered to be indicating deception (NITV, 2015). If no questions are classified as deceptive, but any relevant questions are classified as having a tendency toward deception, a third test is administered (NITV, 2015). If the third test yields deceptive results, the test and subject are considered to be indicating deception (NITV, 2015). If the third test is classified as non-deceptive or having a tendency toward deception, the test and subject are considered non-deceptive (NITV, 2015).

The NITV's Operating Manual (2015) for CVSA cites a study entitled Field Evaluation of Effectiveness of VSA (Voice Stress Analysis) Technology in a US Criminal Justice Setting. That study, conducted by Professor James Chapman of State University of New York at Corning, looked at 236 cases in which a CVSA was used (NITV, 2015 p. 145). 91% of these cases had reached an "investigative impasse" (NITV, 2015 p. 145). According to this study the CVSA had "an accuracy rate of 99.69%, a precision rate of 99.67%, and a verified confession rate of 96.4%" (NITV, 2015 p. 145).

However, other studies did not report similar results. In a 2007 study, researchers conducted voice stress analysis exams on inmates at the Oklahoma County Detention Center (Damphousse et al., 2007). The issue addressed by the exams was whether or not inmates had recently used drugs (Damphousse et al., 2007). After the voice stress analysis exams were conducted, the inmates submitted to drug tests to determine whether or not they had, in fact, recently used drugs (Damphousse et al., 2007). The researchers then had verifiable evidence that could be compared to the voice stress analysis results. They tested the NITV's CVSA system as well as another voice stress analysis system, Layered Voice Analysis (LVA)

(Damphousse et al., 2007). From their conclusion:

Both programs failed consistently to correctly identify respondents who were being deceptive. On average, only about 15% of the respondents who recently used drugs but reported that they had not used drugs were identified as being deceptive. For CVSA, this figure was 8%... (Damphouse et al., 2007, p. 88)

They go on to say, "All previously published research conducted in a lab setting has failed to find support for VSA theory or technology" followed by citations from ten other studies between 1979 and 1996 (Damphousse et al., 2007, p. 89).

Because there is no scientific consensus on the validity of the theory or testing process behind technology-assisted lie detection, the results of an exam are not admissible in court. In *Frye v. United States* (1923), the Supreme Court heard a case in which William Martson's blood pressure deception detection method was used by the defendant, Frye, in an attempt to to exonerate himself for a murder of which he was accused. Frye passed Marston's blood pressure test, but the lower courts did not allow the results of this new technology into evidence (*Frye v. United States*, 1923). The Supreme Court later upheld that ruling and noted:

Just when a scientific principle or discovery crosses the line between the experimental and demonstrable stages is difficult to define...the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs (*Frye v. United States*, 1923, para. 8).

Later, in 1935, a court in Wisconsin allowed the results of a polygraph test administered by Leonarde Keeler to be admitted into evidence (Inbaut, 1935). This remains an unusual anomaly, but it was a special circumstance in which both the prosecution and defense counsels

agreed to having the results admitted prior to the exam being administered (Inbau, 1935).

More recently, in 1998, the United States Supreme Court affirmed that Polygraph examinations were inadmissible. In *United States v. Scheffer*, Justice Thomas wrote, "There is simply no consensus that polygraph evidence is reliable" and upheld the exclusion of polygraph results in the case of a military court-martial (para. 4).

While most rulings regarding lie detector tests have focused on the Polygraph, voice-based lie detection technologies have also been addressed by the courts. In *Smith v. State*, out of the Appellate Court of Maryland from 1976, the court looked at a case involving Psychological Stress Evaluation (PSE). Judge Mason, delivering the court's decision that the test was inadmissible, stated:

The psychological stress evaluation test is basically a voice lie detector test...The difference, if any, between the psychological stress evaluation test and a lie detector test is too minor and shadowy to justify a departure from our prior decision. A lie detector test by any other name is still a lie detector test (*Smith v. State*, 1976, Section V, para. 7).

This ruling, along with the Frye ruling and others, has thus far connected the fate of voice-based lie detection to that of the Polygraph test: inadmissible in court.

Despite the lack of scientific consensus and inadmissibility in court proceedings, CVSA systems are still widely used by law enforcement in the United States. According to the NITV, their CVSA platform is used by over 2,500 law enforcement agencies in the United States, including 41 agencies in Arkansas (NITV, 2020a). The saving grace of the CVSA process is the fact that, though the actual test results are inadmissible, any confessions made during the

process are admissible, so long as other applicable laws, such as Miranda, are followed appropriately. In 2010 the Appellate Division of the New York Supreme Court ruled in *People v. Pollard* that "the use of the CVSA test as an interview tool did not provoke an involuntary confession" and was admissible. In that case, the conducting of a CVSA exam by police was never mentioned at trial, but the admissions from Pollard that came from the CVSA interviews were admitted (*People v. Pollard*, 2010). Pollard argued that the fact that his admissions were in the context of a CVSA exam was exculpatory, and that the omission of that fact at trial meant his admissions should have been excluded (*People v. Pollard*, 2010). However, the court ruled that "The use of the CVSA test as an interview tool did not constitute exculpatory evidence and was not necessary to provide a complete narration of defendant's inculpatory statements" (*People v. Pollard*, 2010).

Based on these rulings, the true utility in voice stress analysis as a means of detecting deception does not lie in the procedure's ability to produce charts for evaluation and presentation in court. Rather, the usefulness of the tool comes from the psychological advantage of the test subject's perception of the procedure's ability to determine truth from lies. When used properly by a trained examiner who understands the psychology of convincing a subject that the truth will be revealed through the procedure, the exam becomes a means to an end - a confession - rather than the end itself.

A similar situation can be found when it comes to the use of Portable Breath Testing (PBT) devices. Under Arkansas law, the only breath test results that are admissible in court are those obtained from devices approved for such use by the Arkansas Department of Health (Arkansas Department of Health, 2014). No PBT devices meet that standard, and thus none are

admissible in court. However, many departments have and use PBT devices during DWI enforcement. While their results are inadmissible in court, officers can use them on suspected drunk drivers or minors who are believed to have been drinking, and those officers can take the results into consideration as they proceed with their investigations. Officers can also use in court any admissions or utterances made by suspects during interviews conducted in conjunction with the PBT testing. In much the same way, CVSA examinations can be a valuable tool as investigators make decisions on how to proceed with their investigations.

While it is a powerful tool for law enforcement, that power can be used for evil just as it can be used for good. Take the case of Michael Crowe, Aaron Houser, and Joshua Treadway out of Escondido, California (Crowe v. County of San Diego, 2010). In 1998 Michael's sister, Stephanie Crowe, was brutally stabbed to death in their home (Crowe v. County of San Diego, 2010). Stephanie was twelve years old at the time, and Michael was fourteen (Crowe v. County of San Diego, 2010). Michael and his family were brought in for questioning, since they were in the house at the time of the murder (Crowe v. County of San Diego, 2010). Investigators keyed in on what they felt was a discrepancy in Michael's description of his activities that evening (Crowe v. County of San Diego, 2010). He said he had gotten out of bed around 4:30 in the morning to get headache medicine from the kitchen (Crowe v. County of San Diego, 2010). He told investigators that all of the bedroom doors he had passed were closed, but investigators believed that Stephanie was already dead by time (Crowe v. County of San Diego, 2010). She was found in the doorway to her room the next morning, so investigators did not believe Michael's account of her bedroom door being closed (Crowe v. County of San Diego, 2010).

Police investigators interviewed or interrogated Michael four times in total (Crowe v.

County of San Diego, 2010). During the third interview, they asked Michael to take a CVSA exam (Crowe v. County of San Diego, 2010). He agreed, but also expressed some frustration because he felt like he was being blamed for his sister's death (Crowe v. County of San Diego, 2010). He took the exam, and the examiner told him he was showing deception (Crowe v. County of San Diego, 2010). Michael maintained his innocence for some time (Crowe v. County of San Diego, 2010). However, the investigators began telling him they knew he did it, and they eventually introduced the theory that there might be another version of him that he was not aware of and that the second version of Michael may have killed Stephanie (Crowe v. County of San Diego, 2010). They began asking him questions that assumed his guilt, such as "can you tell me what you did with the knife?" (Crowe v. County of San Diego, 2010). Michael continued to deny any involvement in his sister's death, but the investigators persisted (Crowe v. County of San Diego, 2010). Under the repeated accusations and suggestion that there may be a "second" Michael, Michael eventually went from strong denials to saying he did not remember killing Stephanie (Crowe v. County of San Diego, 2010).

Michael did not admit to anything during the third interview (*Crowe v. County of San Diego*, 2010). However, during the six hour long fourth interview, investigators began telling him they had physical evidence proving he killed Stephanie (*Crowe v. County of San Diego*, 2010). They demanded he explain evidence they claimed to have, telling him he was not allowed to answer with "I don't know" (*Crowe v. County of San Diego*, 2010). Eventually, Michael said he would tell them a story, but repeatedly told them it was a lie (*Crowe v. County of San Diego*, 2010). He gave them a vague description of how he killed his sister, and they took that as a confession (*Crowe v. County of San Diego*, 2010). They tried to get him to give more

details that could be verified, but he was unable to do so (Crowe v. County of San Diego, 2010).

Investigators decided that Michael's friends, Aaron and Joshua, were also involved in the killing (*Crowe v. County of San Diego*, 2010). They employed similar tactics to try to get them to confess, including twelve and fourteen hour interviews and a CVSA examination where they were told they had shown deception (*Crowe v. County of San Diego*, 2010). Aaron maintained his innocence the entire time, but investigators eventually got Joshua to give a confession (*Crowe v. County of San Diego*, 2010). However, his confession was weak and contained details that were verifiably untrue (*Crowe v. County of San Diego*, 2010).

All three boys were arrested and charged for the murder of Stephanie Crowe (*Crowe v. County of San Diego*, 2010). A judge would later throw out significant portions of the boys' interviews, stating that those portions were coerced and therefore inadmissible (*Crowe v. County of San Diego*, 2010). However, the case was still set for trial until forensics revealed a new suspect (*Crowe v. County of San Diego*, 2010). That man, Richard Tuite, had been seen in the area the day of the murder acting strange and harassing people in the neighborhood (*Crowe v. County of San Diego*, 2010). He was questioned initially, but released (*Crowe v. County of San Diego*, 2010). His shirt was collected, and when it was finally tested, blood was found on the shirt (*Crowe v. County of San Diego*, 2010). That blood was positively matched with Stephanie's DNA (*Crowe v. County of San Diego*, 2010). The case against the boys was quickly dropped. Tuite was charged with voluntary manslaughter and convicted at a jury trial in 2004 (*Crowe v. County of San Diego*, 2010).

The boys and their families sued the City of Escondido, the individual investigators, and the NITV for the role their CVSA played in the investigation and the boys' arrests (*Crowe v*.

County of San Diego, 2010). The city paid a settlement that was publicly released: \$7.25 million (Sharma & Sauer, 2011). The NITV also paid a settlement, but the amount was kept private and they did not admit any fault (Marshall, 2005). A spokesperson for NITV said they stand by their product (Marshall, 2005).

The Crowe case is horrific and an example of what can happen when the CVSA, or any technology-assisted lie detection device, is used maliciously in combination with other inappropriate investigative techniques, like extended interrogation of minors. However, while malice and intentional malpractice can be easier to identify due to their egregious nature, the misuse of the CVSA by well-meaning but incompetent examiners can more easily go unnoticed, but still have a significantly detrimental impact to case investigation. As covered earlier in this paper, the CVSA manual states that failure to properly construct questions can lead to inaccurate results. Imagine a homicide investigation in which a suspect is being given the CVSA examination. An unskilled examiner may be interviewing an innocent person, but improper formulation of irrelevant questions may elicit a stress response unrelated to the relevant question preceding it. This can result in what appears to be delayed stress from a relevant question, when it is actually stress from an improperly formed irrelevant question. Consider the question "Did you kill Jane Doe?" followed by "Are you married?" The first question is a well formed relevant question. It is specific, simple, and gets to the issue at hand. An innocent person should show no significant stress on this question. However, if that innocent person is cheating on their spouse, the follow-up question about their marital status may very well cause stress. Since the examiner is unaware of the stress related to the marriage question, they may interpret that stress as delayed stress from the relevant question. This may convince the

investigator that they have a good suspect, which may delay or even prevent the investigation into and identification of the true murderer.

Other problems can arise at various stages of the examination process. An examiner who fails to adequately explain the process may not convince a guilty subject of its legitimacy, which may greatly reduce the subject's stress and cause them not to show significant enough deception when lying to fail the test. There could also be an examiner who starts with a hardline, accusatory interrogation just before the exam, which may put an innocent person in such a heightened state of stress that they show deception even when telling the truth. There could also be situations where an examiner who does not fully explain the reasoning for the second chart may lead to an innocent subject thinking they failed the first chart, thus increasing their overall stress and leading to a chart that indicates deception. All of these issues can occur even when the examiner has nothing but the best intentions.

Problems of bad-faith investigators misusing the CVSA can only be solved with departmental or external oversight. However, problems of incompetence may be addressed through training. Currently, the NITV only requires a 5 day course to become a certified examiner (NITV, 2020b). While this course does incorporate some practical exercises, the certification exam is only a written test. There is no requirement that students demonstrate proficiency with the equipment or in performing complete examinations. Recertification is only required every two years, and there is no minimum number of actual examinations required over that time period in order to maintain certification status. The performance of actual examinations requires knowledge of equipment, interview skills, highly-specific question structuring, the ability to modify questions based on feedback from test subjects, strict

adherence to testing protocols, and interpretation of scientific charts. A deficiency in any of these areas can put the process at risk. Because there is only one week of training for certification, and because there is no practical examination requiring a display of proficiency, and because examiners only have to recertify every two years, and because there is no minimum number of tests that must be performed in two years, there can be a great variation in the actual ability of different examiners in the field. An examiner from a small department may only conduct one or two exams per year, whereas an examiner at a larger department could conduct dozens of exams in the same time frame. This disparity does nothing but hurt the credibility of the technology, which will likely guarantee it maintains its status as a useful tool for examiners dedicated to the craft, a stumbling block for the well-meaning but unmotivated, a dangerous weapon for those who would intentionally misuse it, but inevitably inadmissible in court for everyone. The Computerized Voice Stress Analyzer has a place in law enforcement, but like all tools, its value rests entirely in the person wielding it.

References

Arkansas Department of Health (March 2014). *Arkansas breath testing operator training manual*.

https://www.healthy.arkansas.gov/images/uploads/pdf/OperatorTrainingManual.pdf
Crowe v. County of San Diego, 593 F.3d 841. 9th Cir. (2010)

Damphousse, K., Pointon, L., Upchurch, D. & Moore, R (2007, March 31). *Assessing the validity of voice stress analysis tools in a jail setting.* Research Report. (National Institute of Justice Grant 2005-IJ-CX-0047)

https://www.ncjrs.gov/pdffiles1/nij/grants/219031.pdf

Frye v. United States, 293 F. 1013. D.C. Cir. (1923)

Inbau, F (1935). *Detection of deception technique admitted as evidence*. Journal of Criminal Law and Criminology.

https://scholarlycommons.law.northwestern.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=2559&context=jclc

Marshall, S. (2005, May 25). *Maker of voice-analysis machine settles Crowe lawsuit*. San Diego Union Tribune. https://www.sandiegouniontribune.com/sdut-maker-of-voice-analysis-machine-settles-crowe-2005may25-story.html

National Research Council, *The polygraph and lie detection*. (Washington, D.C.: The National Academies Press, 2003.

NITV Federal Services. Certified examiners course operating manual. (2015)

NITV Federal Services (2017). A guide to formulating interview questions for a computer voice stress analysis exam. https://www.cvsa1.com/interviewing-and-interrogation/a-guide-to-formulating-interview-questions-for-a-computer-voice-stress-analysis-exam/

NITV Federal Services (2020a). Agencies using CVSA in the USA.

https://www.cvsa1.com/agencies.htm

NITV Federal Services (2020b). Certified examiner's course - 2020 class schedule.

https://www.cvsa1.com/cecschedule.htm

People v. Pollard, 70 A.D.3d 1403. 894 N.Y.S.2d 691 (2010)

Sharma, A. & Sauer, M. (2011, October 2). *\$7.25 million settlement reached in Stephanie Crowe murder case*. KPBS. https://www.kpbs.org/news/2011/oct/21/7-million-settlement-reached-stephanie-crowe-murde/

Smith v. State, 355 A.2d 527. 31 MD App. 106 (1976)

United States v. Scheffer, 96-1133. 523 U.S. 303 (1998)