



Voice Stress Analysis: Only 15 Percent of Lies About Drug Use Detected in Field Test

by Kelly R. Damphousse, Ph.D.

Law enforcement agencies across the country have invested millions of dollars in voice stress analysis (VSA) software programs.¹ One crucial question, however, remains unanswered:

Does VSA actually work?

According to a recent study funded by the National Institute of Justice (NIJ), two of the most popular VSA programs in use by police departments across the country are no better than flipping a coin when it comes to detecting deception regarding recent drug use. The study's findings also noted, however, that the mere presence of a VSA program during an interrogation may deter a respondent from giving a false answer.

VSA manufacturers tout the technology as a way for law enforcers to accurately, cheaply, and efficiently determine whether a person is lying by analyzing changes in their voice patterns. Indeed, according to one manufacturer, more than 1,400 law enforcement

agencies in the United States use its product.² But few studies have been conducted on the effectiveness of VSA software in general, and until now, none of these tested VSA in the field—that is, in a real-world environment such as a jail. Therefore, to help determine whether VSA is a reliable technology, NIJ funded a field evaluation of two programs: Computer Voice Stress Analyzer[®] (CVSA[®])³ and Layered Voice Analysis[™] (LVA).

Researchers with the Oklahoma Department of Mental Health and Substance Abuse Services (including this author) used these VSA programs while questioning more than 300 arrestees about their recent drug use. The results of the VSA output—which ostensibly indicated whether the arrestees were lying or telling the truth—were then compared to their urine drug test results. The findings of our study revealed:

- **Deceptive respondents.** Fifteen percent who said they had not used drugs—but

who, according to their urine tests, had—were *correctly* identified by the VSA programs as being deceptive.

- **Nondeceptive respondents.** Eight and a half percent who were telling the truth—that is, their urine tests were consistent with their statements that they had or had not used drugs—were *incorrectly* classified by the VSA programs as being deceptive.

Using these percentages to determine the overall accuracy rates of the two VSA programs, we found that their ability to accurately detect deception about recent drug use was about 50 percent.

Based solely on these statistics, it seems reasonable to conclude that these VSA programs were not able to detect deception about drug use, at least to a degree that law enforcement professionals would require—particularly when weighed against the financial investment. We did find, however, that arrestees who were questioned using the VSA instruments were less likely to lie about illicit drug use compared to arrestees whose responses were recorded by the interviewer with pen and paper.

So perhaps the answer to the question “Does VSA work?” is . . . it depends on the definition of “work.”

What Is VSA?

VSA software programs are designed to measure changes in voice patterns caused by the stress, or the physical effort, of trying to hide deceptive responses.⁴ VSA programs interpret changes in vocal patterns and indicate on a graph whether the subject is being “deceptive” or “truthful.”

Most VSA developers and manufacturers do not claim that their devices detect lies; rather, they claim that VSA detects micro-tremors, which are caused by the stress of trying to conceal or deceive.

VSA proponents often compare the technology to polygraph testing, which attempts to measure changes in respiration, heart rate, and galvanic skin response.

Even advocates of polygraph testing, however, acknowledge its limitations, including that it is inadmissible as evidence in a court of law; requires a large investment of resources; and takes several hours to perform, with the subject connected to a machine. Furthermore, a polygraph cannot test audio or video recordings, or statements made either over a telephone or in a remote setting (that is, away from a formal interrogation room), such as at an airport ticket counter. Such limitations of the polygraph—along with technological advances—prompted the development of VSA software.

Out of the Lab, Into the Field

Although some research studies have shown that several features of speech pattern differ under stress,^{5,6} it is unclear whether VSA can detect *deception-related* stress. In those studies that found that this stress *may* be detectable, the deception was relatively minor and no “jeopardy” was involved—that is, the subjects had nothing to lose by lying (or by telling the truth, for that matter). This led some researchers to suggest that if there is no jeopardy, there is no stress—and that if there is no stress, the VSA technology may not have been tested appropriately.⁷

The NIJ-funded study was designed to address these criticisms by testing VSA in a setting where police interviews commonly occur (a jail) and asking arrestees about relevant criminal behavior (drug use) that they would likely hide.⁸

Our research team interviewed a random sample of 319 recent arrestees in the Oklahoma County jail. The interviews were conducted in a relatively private room adjacent to the booking facility with male arrestees who had been in the detention facility for less than 24 hours. During separate testing periods, data were collected using CVSA[®] and LVA.

The arrestees were asked to respond to questions about marijuana use during the previous 30 days, and cocaine, heroin, methamphetamine, and PCP use within the previous 72 hours. The questions and

Editor's Note

POLYGRAPH AND VOICE STRESS ANALYSIS: TRYING TO FIND THE RIGHT TOOL

The validity of the polygraph as a lie-detection device has been under fire for years. In 2003, the National Academy of Sciences issued a report identifying major deficiencies in polygraph technology.⁹ The report and other analyses led to the research and development of potential alternatives to the polygraph; one technology that emerged is voice stress analysis (VSA).

The National Institute of Justice funded a study to evaluate two of the most popular VSA software programs in a real-world (that is, nonlaboratory) setting in which jeopardy—the threat of penalty—was present.

The study found that the average accuracy rate of these programs in detecting deception regarding drug use was approximately 50 percent—about as accurate as flipping a coin. But the research also found that subjects may be deterred from lying if they think their responses can be “proven” false.

It remains to be seen, however, if any deterrence factor dissipates as word spreads about the accuracy rate of VSA software programs. Prospective users of VSA should weigh all these factors, including that there may be an investigative, even if there is no evidentiary, use for this technology.

test formats were approved by officials from CVSA[®] and LVA. The VSA data were independently interpreted by the research team and by certified examiners from both companies.

Following each interview, the arrestee provided a urine sample that was later tested for the presence of the five drugs. The results of the urinalysis were compared to the responses about recent drug use to determine whether the arrestee was being truthful or deceptive. This determination was then compared to the VSA output results to see whether the VSA gave the same result of truthfulness or deceptiveness.

Can VSA Accurately Detect Deception?

Our findings suggest that these VSA software programs were no better in determining deception about recent drug use among arrestees than flipping a coin.

To arrive at this conclusion, we first calculated two percentage rates¹⁰:

- **Sensitivity rate.** The percentage of deceptive arrestees correctly identified by the VSA devices as deceptive.
- **Specificity rate.** The percentage of non-deceptive arrestees correctly classified by the VSA as nondeceptive.

Both VSA programs had a low sensitivity rate, identifying an average of 15 percent of the responses by arrestees who lied (based on the urine test) about recent drug use for all five drugs. LVA correctly identified 21 percent of the deceptive responses as deceptive; CVSA[®] identified 8 percent.

The specificity rates—the percentage of nondeceptive respondents who, based on their urine tests, were correctly classified as nondeceptive—were much higher, with an average of 91.5-percent accuracy for the five drugs. Again, LVA performed better, correctly identifying 95 percent of the nondeceptive respondents; CVSA[®] correctly identified 90 percent of the nondeceptive respondents.

We then used a plotting algorithm, comparing the sensitivity and specificity rates, to calculate each VSA program’s overall “accuracy rate” in detecting deception about drug use.¹¹ We found that the average accuracy rate for all five drugs was approximately 50 percent.

Does VSA Deter People From Lying?

Although the two VSA programs we tested had about a 50-percent accuracy rate in determining deception about recent drug use, might their very presence during an interrogation compel a person to be more truthful?

This phenomenon—that people will answer more honestly if they believe that their responses can be tested for accuracy—is called the “bogus pipeline” effect.¹² Previous research has established that it is often present in studies that examine substance use.¹³

To determine whether a bogus pipeline effect existed in our study, we compared the percentage of deceptive answers to data from the Oklahoma City Arrestee Drug Abuse Monitoring (ADAM) study (1998–2004), which was conducted by the same VSA researchers in the same jail using the same protocols. The only differences—apart from the different groups of arrestees—were that the ADAM survey was longer (a 20-minute survey compared with the VSA study’s 5-minute survey) and did not involve the use of VSA technology.

In both studies, arrestees were told that they would be asked to submit a urine sample after answering questions about their recent drug use. In the VSA study, arrestees were told that a computer program was being used that would detect deceptive answers.

Arrestees in the VSA study were much less deceptive than ADAM arrestees, based on responses and results of the urine test (that is, not considering the VSA data). Only 14 percent of the VSA study arrestees were deceptive about recent drug use compared to 40 percent of the ADAM arrestees. This suggests that the arrestees in the VSA study who thought their interviewers were using a form of “lie detection” (i.e., the VSA technology) were much less likely to be deceptive when reporting recent drug use. (See sidebar on p. 10, “Editor’s Note, Polygraph and Voice Stress Analysis: Trying to Find the Right Tool.”)

The Bottom Line: To Use or Not Use VSA

It is important to look at both “hard” and “hidden” costs when deciding whether to purchase or maintain a VSA program. The monetary costs are substantial: it can cost up to \$20,000 to purchase LVA. The average cost of CVSA[®] training and equipment is \$11,500. Calculating the current investment nationwide—more than 1,400 police departments currently use CVSA[®], according to the manufacturer—the total cost is more than \$16 million not including the manpower expense to use it.

The hidden costs are, of course, more difficult to quantify. As VSA programs come under greater scrutiny—due, in part, to reports of false confessions during investigations that used VSA—the overall value of the technology continues to be questioned.¹⁴

Therefore, it is not a simple task to answer the question: Does VSA work? As our findings revealed, the two VSA programs that we tested had approximately a 50-percent accuracy rate in detecting deception about drug use in a field (i.e., jail) environment; however, the mere presence of a VSA program during an interrogation may deter a respondent from answering falsely. Clearly, law enforcement administrators and policy-makers should weigh all the factors when deciding to purchase or use VSA technology.

NCJ 221502

Notes

1. The National Institute for Truth Verification (manufacturer of CVSA[®]) states that more than 1,400 law enforcement agencies use its product. See www.nitv1.com/Agenciesusing.htm.
2. Ibid.
3. CVSA[®] was introduced into the market in 1988 by the National Institute for Truth Verification and has undergone a number of changes and system upgrades over the years. The version used in this field test was the CVSA[®] introduced in 1997.
4. Hopkins, C.S., R.J. Ratley, D.S. Benincasa, and J. Grieco, “Evaluation of Voice Stress Analysis Technology,” *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, 2005.
5. In the few studies in which the theory behind VSA has been tested, there has generally been solid support. Cestaro, V.L., “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) (1996): 117–127; and Fuller,

The products, manufacturers, and organizations discussed in this document are presented for informational purposes only and do not constitute product approval or endorsement by the U.S. Department of Justice.

- B.F., "Reliability and Validity of an Interval Measure of Vocal Stress," *Psychological Medicine* 14 (1) (1984): 159–166.
6. Researchers at the Air Force Research Laboratory concluded that two VSA devices (Lantern™ and the Psychological Stress Evaluator—a precursor of CVSA®) could measure these differences in speech patterns. Hansen, J., and G. Zhou, *Methods for Voice Stress Analysis and Classification: Final Technical Report*, Rome, NY: U.S. Air Force Research Laboratory, 1999; and Haddad, D., S. Walter, R. Ratley, and M. Smith, *Investigation and Evaluation of Voice Stress Analysis Technology*, final report submitted to the National Institute of Justice, 2002 (NCJ 193832), available at www.ncjrs.gov/pdffiles1/nij/193832.pdf.
 7. Barland, G., "The Use of Voice Changes in the Detection of Deception," *Polygraph* 31 (2) (2002): 145–153. This study suggests simulated stress in a laboratory setting may not be sufficient to allow VSA to detect deception. This leads to the argument, by some VSA proponents, that mock deception in a staged (lab) scenario fails to create the necessary degree of jeopardy (and therefore stress) to stimulate a measurable response indicating deception. In an experiment in which the subject is not worried about getting "caught" because there are no real consequences or is pretending to lie, it is, they argue, more difficult for the software to detect deception, as the necessary stress levels are not present.
 8. Previous arrestee studies suggest that respondents are commonly deceptive about recent drug use. Fendrich, M., and Y. Xu, "Validity of Drug Use Reports from Juvenile Arrestees," *International Journal of the Addictions* 29 (8) (1994): 971–985; Hser, Y.I., "Self-Reported Drug Use: Results of Selected Empirical Investigations of Validity," *NIDA Research Monograph* 167 (1997): 320–343; Lu, N.T., B.J. Taylor, and K.G. Riley, "The Validity of Adult Arrestee Self-Reports of Crack Cocaine," *American Journal of Drug and Alcohol Abuse* 27 (3) (2000): 399–407; Mieczkowski, T., D. Barzelay, B. Gropper, and E. Wish, "Concordance of Three Measures of Cocaine Use in an Arrestee Population: Hair, Urine, and Self-Report," *Journal of Psychoactive Drugs* 23 (3) (1991): 241–249; and Harrison, L., "The Validity of Self-Reported Data on Drug Use," *Journal of Drug Issues* 25 (1) (1995): 91–111.
 9. Committee to Review the Scientific Evidence on the Polygraph, National Research Council, *The Polygraph and Lie Detection*, Washington, DC: National Academies Press, 2003.
 10. Ibid.
 11. Sensitivity and specificity should be examined jointly, because an overly sensitive but not specific instrument—that is, one that indicates all responses as deceptive—is not very useful. The standard way to compare these two scores simultaneously is by examining them on a receiver operating characteristic chart. Programs with high sensitivity and specificity scores will efficiently predict who is being deceptive and who is not. If either the sensitivity or the specificity score is low, the usefulness of the programs for predicting deception is diminished.
 12. Jones, E.E., and H. Sigall, "The Bogus Pipeline: A New Paradigm for Measuring Affect and Attitude," *Psychological Bulletin* 76 (5) (1971): 349–364.
 13. Aguinis, H., C.A. Pierce, and B.M. Quigley, "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) (1993): 352–373; Botvin, E.M., G.J. Botvin, N.L. Renick, A.D. Filazzola, and J.P. Allegrante, "Adolescents' Self-Reports of Tobacco, Alcohol, and Marijuana Use: Examining the Comparability of Video Tape, Cartoon and Verbal Bogus-Pipeline Procedures," *Psychological Reports* 55 (1984): 379–386; and Sprangers, M., and J. Hoogstraten, "Response-Style Effects, Response-Shift Bias and a Bogus-Pipeline," *Psychological Reports* 61 (1987): 579–585.
 14. Hansen, M., "Untrue Confessions," *ABA Journal*, July 1999, 50–53; Wagner, D., "Arguments Rage Over Voice-Stress Lie Detector," *Arizona Republic*, October 10, 2005; and "Innocent Until Proved Guilty?" *ABC News*, March 30, 2006, available at www.abcnews.go.com/Primetime/story?id=1786421&page=1.

About the Author

Kelly Damphousse is associate dean of the College of Arts and Sciences and Presidential Professor of Sociology at the University of Oklahoma. He has 20 years of criminal justice and drug research experience. From 1998 to 2004, Damphousse served as the site director of the Arrestee Drug Abuse Monitoring program in Oklahoma City and Tulsa, Oklahoma. He has directed several statewide and nationwide program evaluation projects.