



John J. Palmatier, Ph.D.
7619 Southern Brook Bend #203
Tampa, Florida 33635

credibility assessment

The CVSA, Polygraph, and Trustech / Vericator Voice Analysis Technologies



*Preliminary Results from a Comparative
Analysis Conducted in a Criminal Justice
Field Setting*

.....

The CVSA, Polygraph, and Trustech / Vericator Voice Analysis Technologies

Preliminary Results from a Comparative Analysis Conducted in a Criminal Justice Field Setting

Introduction

Recently, in a technical report (Palmatier, 2000) the author wrote that: law enforcement agencies across the United States have long relied on the polygraph as one of many tools used to facilitate their ability to serve, protect, and defend their respective communities. However, the prohibitions mandated by the federally enacted Employee Polygraph Protection Act of 1988, and budgetary restraints coupled with equipment and training costs, have severely limited the availability of polygraph resources in many states. Today, relatively few local, county or municipal, law enforcement agencies employ even a single polygraph examiner. Hence, state police and other law enforcement agencies with polygraph units often incur prohibitively large backlogs, necessitating that they limit the types of cases they will accept for examination. Consequently, many law enforcement agencies do not have access to these services, or at best may only have access to them for the investigation of only the most serious crimes. The net effect for many agencies is that a large void exists between their desire to conduct professional investigations and their ability to assess suspect, witness and, too often, victim credibility.

Today, law enforcement agencies primarily use one of three technologies to assess credibility. The oldest, and most researched, is the polygraph, more recently referred to as the psycho-physiological detection of deception (PDD). The validity of the polygraph procedure as it is used in criminal justice settings is well documented with a significant amount of the literature attesting to the procedure's validity (e.g., Barland, 1988; Podlesny & Raskin, 1978).

Chronologically, the second technology is voice stress analysis (VSA), which was developed in the mid 1960's as a result of United States Army research conducted in an attempt to find a less invasive way of detecting deception. According to a former Director of the United States Army Polygraph School (R. Decker, personal conversation, May 4, 1999), the VSA research was unsuccessful and the technology never adopted by the military.

The military VSA project manager, and one of the engineers involved in that research, retired leaving government service. Together they created and marketed the first commercial VSA instrument called a Psychological Stress Evaluator (PSE).

•
•
•
•
•
•
•
•



Charles Humble, a former PSE operator and salesman, later used the PSE technology to have developed another VSA device that he described as an improved instrument for the detection of deception (C. Humble, personal communication, November 1, 1995). In 1988 Humble, doing business as the National Institute for Truth Verification (NITV), began marketing what he described as a Computerized Voice Stress Analyzer (CVSA).

The NITV now claims that more than 900 law enforcement agencies have purchased and/or use the CVSA instrument. In its marketing literature, the NITV also reports impressive statistics regarding the CVSA's accuracy (98%) and the number of confessions obtained using the instrument. However, questions regarding the CVSA's validity in a criminal justice setting, that is can a law enforcement examiner using a CVSA instrument accurately discriminate between truthful and deceptive people; and questions regarding its reliability, that is, are the CVSA's results consistent from one period to another, are issues not yet addressed in a field setting (Palmatier, 1999).

The third technology used to assess credibility is voice analysis (VA), which resulted from efforts made by Amir Liberman, an Israeli computer programmer. In the mid 1990s, due in part to an increase in terrorist activities within Israel, Liberman became increasingly concerned about the difficulty in properly screening people entering his home country. Liberman said (personal conversation, February 3, 2000) he tried to develop software using VSA technology, but was unable to accurately discriminate between truthful and deceptive subjects. Liberman discussed his ideas and the problems he encountered with some academics from the Hebrew University in Jerusalem. After many discussions a basic algorithm evolved. The algorithm focused on several new parameters for the assessment of credibility including the analysis of the entire vocal spectrum and a comparative analysis of individual differences in some personal speech characteristics. Liberman used this algorithm as a basis for his first software program, which showed promise in developmental testing. These results were then used to create interest in the technology in Israel and as a base on which the Trustech organization evolved. After approximately three years of development a professional version of Liberman's software was marketed and has sold in several different countries around the globe (for a more detailed explanation of the software the reader is directed to patent PCT/IL98 /00613, International publication number WO 99/31653). The software using this technology in United States is called Vericator, and for the rest of this paper Vericator and Trustech will be used synonymously.

Differences in the Technologies. Each of the technologies is unique given the manner in which credibility is assessed from the data used. The polygraph process uses physical attachments to record an individual's respiration, electrical potential of the skin, and cardiovascular activity. These parameters are continually monitored as a person is asked relevant questions, which address the issue at examination; control questions, which a subject will have some doubt or will purposely lie to when answering; and irrelevant questions, which are used to establish a baseline for each person examined (Horvath, 1988). The polygraph instrument creates tracings for each channel recorded, which are then analyzed by a trained examiner who must interpret them and render an opinion as to the tested person's credibility.

•
•
•
•
•
•
•
•

Unlike the polygraph, the CVSA (VSA technology) and the Vericator (VA technology) software programs do not require any physical attachments to be placed on the person being examined. However, in order to create the voice tracings for analysis, the CVSA procedure requires the use of the same types of questions used by polygraph examiners. Like the charts produced in a polygraph examination, the CVSA voice data recordings must be analyzed and interpreted by a trained examiner who will then render an opinion regarding a person's credibility. The CVSA tracings are a result of the software program filtering out all but the 8 to 14 Hz portion of the frequency spectrum, and then displaying changes in both AM (audio modulated) and FM (frequency modulated) activity.

The Vericator software program differs from the CVSA in that it is much more flexible and does not require the use, per se, of relevant, control, and irrelevant questions. The Vericator software can be used as easily to assess free-flowing conversation as it can to examine a structured question format like that used in a polygraph or CVSA examination. The Vericator software also differs from the CVSA program by analyzing the entire vocal spectrum including high, medium, and low frequencies. Additionally, the Vericator software uses individual differences in some speech characteristics to aid the examiner in formulating an opinion regarding a person's credibility.

Purpose of this study. The differences in the measures recorded using a polygraph, the CVSA or the Vericator procedures may or may not contribute to differences in both the validity and the reliability of these technologies. Prior to this study, the comparative accuracy of these three technologies had never been explored in the same context. This study was conducted, in part, to address this issue.

Methods

Subjects

The sample consisted of 77 subjects drawn from one of two groups. Subjects in the two groups were drawn from a pool of 296 consecutive field CVSA/polygraph examinations given between July 2, 1997, and May 13, 1998, at a state police polygraph unit during the course of different criminal investigations conducted by township, city, county, state and federal law enforcement agencies. Subjects in the first group (n=36) were confirmed deceptive by giving a confession outlining details relevant to the offense under investigation, while subjects in the second group (n=41) were confirmed truthful as someone else confessed to the crime for which each subject was tested. The crimes committed by the subjects who were deceptive included murder, rape, robbery, larceny, and fraud.

Apparatus

The polygraph examinations and voice recordings were conducted in a small, quiet room used for field examinations (Reid & Inbau, 1977). Subjects sat in a standard Lafayette polygraph chair with adjustable arm rests. The recording of all physiological data, excluding the voice, was accomplished through the use a computerized polygraph instrument. The instrument was a Lafayette Model LX-

•
•
•
•
•
•
•
•
•
•

3000W system manufactured by the Lafayette Instrument Company of Lafayette, Indiana. The computer using the polygraph software was a Dell desktop computer with 64 MB of RAM memory using an Intel Pentium processor operating at 266 MHz. Thoracic and abdominal respiration was recorded for each subject by means of two pneumograph tubes, one positioned around the thorax and the other around the abdomen. Skin resistance response (SRR) was recorded through two stainless steel electrodes attached to the volar surface of the first and second fingers of either hand. No electrode gel was used for these examinations; the electrodes were dry as is the custom in field examinations. Cardiovascular activity was recorded by use of a standard pneumatic pressure cuff around each subject's arm. Cuff pressure was set to about 70 mm/hg and the cuff squeezed two or three times to equalize the air within it, generally resulting in a recording pressure of 50 to 65 mm/hg.

The NITV advocates using live examinations, and the decisions made by the examiners administering them, as the basis on which the CVSA's validity should be assessed and does not support the use of tape recordings for this purpose. However, there were several problems associated with using live CVSA examinations that would make evaluating the scientific validity of the CVSA difficult, if not impossible.

For instance, the inter-active nature of live CVSA examinations, like live polygraph examinations, would not permit in any reasonable manner an ability to isolate what the instrument's contribution was to a decision independent of the examiner's observations. The analog CVSA instrument that was used for this study also used a heat sensitive chart paper to record voice data when a subject responds to questions. This heat sensitive chart paper is fragile and makes it a poor choice for permanently archiving data. Further, once an examination was concluded the data could not be reproduced if it was damaged or lost. Finally, while conducting training examinations, it was discovered that the noise created by the analog CVSA instrument while printing a voice chart proved to be a distraction to some examinees in the actual field-testing setting. And the possibility that this distraction could possibly affect an examinee's responses could not be disregarded.

Accordingly, all voice data was recorded at a sampling frequency of 44.1 kHz using a studio grade TASCAM Model DA-20 Digital Audio Tape (DAT) Recorder/Player (TEAC, Montebello, CA). Input and output were through the DAT's left channel analog jacks. The DA-20's input originated with a Model AT803b powered lavalier condenser microphone (Audio-Technica, Stow, OH), which was powered by and amplified through an Aphex Model 107 microphone pre-amplifier (Aphex Systems, Sun Valley, CA). The output from the 107 microphone pre-amplifier to the DAT was through a TRS ¼" phone jack with the selector switch set at the -10dBV position. The DA-20's input (recording) level control, and the 107 microphone pre-amplifier's gain control, were increased from their lowest settings to a level permitting a signal strong enough for recording purposes. For each control this was approximately 20/25% of the controls full range. AMPEX Model 467 R-124 Certified Mastering Audio DAT cassette tapes were used to permanently archive voice data. For playback purposes the DAT's left channel analog output was connected to the input of an Aphex Model 120A distribution amplifier (Aphex Systems, Sun Valley, CA), which was used to properly attenuate the output signal from the DAT Recorder/Player to the CVSA's microphone input. The distribution amplifier's level

•
•
•
•
•
•
•
•

control was adjusted between approximately 10 and 60% of its full range to insure an adequate signal for the CVSA to print voice response charts. The Vericator software does not produce a tracing like the CVSA or polygraph instruments. Therefore, wave files were created using the DAT recordings and the wave files forwarded to the Trustech offices in Israel for analysis.

Examiner

The author served as the primary examiner and was certified by the NITV first in a CVSA Certified Examiners Course (CEC) held at the Cincinnati Police Department in Ohio, in early January 1996. At the request of Charles Humble, at that time CEO for the NITV, a second CEC course was attended at the Palm Beach County Sheriff's Department in West Palm Beach, Florida, in late November 1996, to refresh the author's understanding of VSA procedures and practices. The examiner was also a licensed PDD examiner for the State of Michigan with 24 years of police service and 14 years of experience conducting polygraph examinations in support of criminal investigations. The Royal Canadian Mounted Police at the Canadian Police College in Ottawa, Canada, conducted the examiner's Polygraph training during the fall of 1983.

Procedures

The Examination - Pretest Phase: Subjects are routinely scheduled for polygraph examinations, regarding a variety of crimes ranging from larceny to murder, at one of the polygraph units operated by the Department of State Police within the State of Michigan. To insure an adequate number of participants, subjects for this study were solicited from all examinees arriving for testing at this State Police Polygraph unit. For each examination, administrative personnel notified the examiner when a subject scheduled for testing arrived.

Informed Consent: Each examinee was told that the Department of State Police was continually looking for and testing new technology. The examinee was asked to facilitate this process by giving permission to record their verbal responses to the questions that would be asked later regarding the issue for which they were being tested. The examinee was then given an informed consent form and asked to read it carefully. If the examinee said they wished to participate in the research study they were asked to please sign and date the form (a copy of this form is available from the author).

Advice of Rights and Pretest Questioning: Due to the fact that the examinations were conducted pursuant to a criminal investigation each examinee was advised of his/her Miranda warnings and other rights applicable to the examination setting using a standardized form that is read to all examinees who submit to a polygraph examination conducted by the Department of State Police (a copy of this form is available from the author) . The examinee was then asked why s/he had been scheduled for a polygraph examination? After a short discussion and clarification of the issue to be tested, i.e., larceny, rape, robbery, murder, etc., and the examinee's

•
•
•
•
•
•
•
•
•
•

status (suspect, witness, victim), the examinee was asked a series of routine questions, titled "Personal History" and "Exam Info", included in the Lafayette Instrument's LX3000 computerized polygraph program. Following this, the examinee was given a detailed explanation of the polygraph procedure and the importance of telling the complete truth. The examinee was then asked for a detailed narrative relating his/her knowledge about the issue at examination. With this information, and considering the information given by the investigator, the examination questions were then formulated.

Test Question Format: For each examination there was a total of 4 relevant questions, 2 directed lie control questions and 7 irrelevant questions. For each examination, i.e., for all subjects tested, the only questions that changed were the 4 relevant questions. After all the questions had been formulated they were reviewed with the examinee. The first questions reviewed were the irrelevant questions. Examinees were told that they would be asked some questions that they knew were true and which the examiner knew were true and a "YES" or "NO" answer was needed in response. Each question was asked and the answer noted. Examinees were then read the relevant questions and again a yes or no answer registered. Finally, the control questions were introduced and the examinees instructed to answer these questions with the one word "NO", a lying response.

Research Protocol Endorsement: The order in which questions were asked for testing is described by CVSA advocates as an NITV "General Series" question format (NITV, 1996). Before making a decision to use this question format in this study, it was discussed with the NITV's Chief Instructor, at that time Special Agent Bob Tippett of the Florida Department of Law Enforcement. The proposed "General Series" question format was then included in the research proposal submitted to the NITV for their review prior to submitting the document for funding consideration. The research proposal and the use of the "General Series" question format was then endorsed by David Hughes (a copy of Mr. Hughes endorsement is available from the author), who served at that time as the NITV's Director of Training and is now that organization's Executive Director.

Rationale For Using the NITV Question Format: The NITV's General Series format is different than the General Series format commonly used by polygraph examiners for control question polygraph testing. The two primary differences are related to the number and use of irrelevant questions, and the position in which the first control question is asked. The NITV format asks the first control question in the number 2 position while polygraph tests generally ask the first control question in either position 3, 4, or 6. And second, the NITV format asks an irrelevant question after each control and relevant question. Thereby bracketing the control and irrelevant questions between irrelevant questions to monitor what Charles Humble has termed "delayed stress."

The decision to use the NITV's question format was made to give the CVSA instrument every opportunity to succeed. It is believed that the NITV's "General Series" question format had not been used for polygraph or voice analysis testing anywhere prior to its use for this study. It was also understood that in doing so the polygraph procedure would possibly be placed at a disadvantage. However, research conducted by Horvath (1988) suggested that:

•
•
•
•
•
•
•
•

CO [control question] testing is sufficiently robust that many of the minor differences in the manner in which that testing is structured have little effect (p.208).

Horvath's conclusion was later bolstered by another study (Palmatier, 1991) that tested two common variations of control question tests and found, in part, that:

"the accuracy of the Zone Comparison Test was statistically equivalent to that of the Modified General Question Test" (p. 140).

These findings, coupled with the absence of any contradictory empirical evidence, and the advantage of dispensing with or at least diminishing any arguments that this study's findings could be attributed to the use of dissimilar examination formats, strongly suggested that it was reasonable and prudent to use the NITV's "General Series" question format for both CVSA and polygraph testing.

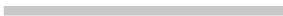
The Examination - Instrumentation Phase: After creating the relevant questions to be asked on each test, the examiner attached the microphone to the front of each participating subject's clothing near the collar. The DAT recorder was then activated.

Collection of Voice Data: Each examinee was instructed to answer all the questions as appropriate remembering to answer the two questions they were directed to lie about (the Directed Lie Controls) with a "NO" answer. The examiner would then read all the questions to the examinee in the prescribed order.

The procedure advocated by the NITV dictates that after reading through the questions once, the examiner should explain to each subject that there are distinct differences in a person's voice between the first and second times they answer a series of questions. Truthful people may experience some apprehension the first time they answer a set of questions, but the fear or apprehension disappears when the questions are answered a second time. Deceptive people, however, only experience-increased apprehension that, although it is not detectable by the human ear, is detectable using proper instrumentation, like that used for the examination. Accordingly, the questions were asked a second time and each subject again asked to answer the questions as appropriate.

Collection of Physiological Data: After recording the questions and answers the second time, the examiners then placed the polygraph attachments on each subject. The examiners first administered an acquaintance/stimulation test and then administered three or more polygraph charts using the questions asked for the two VSA tests. For chart one and chart two, the examiner asked the questions in the same order as that used for the two VSA tests. Chart three also used the same questions, but in a mixed order. The examiners allowed the DAT recorder to continue running during the polygraph procedure. After the polygraph testing was finished, and before the subject was given the examiner's opinion, the attachments were removed. The subject was then told that the examiner wanted to verify the earlier recordings and whether the subject's voice had changed during the course of the examination. The examiner then read through the questions one last time in the same order as that used for the first two VSA tests, and again omitting the twenty-five second delay

•
•
•
•
•
•
•
•



between questions required for the polygraph tests. Once the recording was finished, the tape recording of the examination and a backup copy of the polygraph data were kept in a secure file until a sufficient number of verified cases, truthful and deceptive, were collected.

Data Reduction and Analysis: After the required number of truthful and deceptive cases was verified, copies of the data were created as detailed below and sent to trained CVSA, Vericator, and polygraph evaluators. The evaluator's were asked to render opinions regarding each subject's status using only that information gleaned from the copies of the recorded data they were given. No other information was given to the evaluators regarding the proportion of truthful and deceptive subjects.

The dependent measures for the CVSA, Vericator, and polygraph examinations were the number of correct decisions regarding a subject's status of truthful or deceptive. Although several examiners scored the polygraph and CVSA data, only one examiner analyzed the wave files using the Vericator software. The Vericator software is automated to a degree that a replication of the analyses would show little if any variance. Accordingly, only the results of the most accurate CVSA and polygraph examiners are reported here for comparative purposes. The results for the other examiners also scoring the data will be reported in a more thorough report to be submitted later for publication. The CVSA, Vericator, and the polygraph accuracy rates were determined both by including and excluding the inconclusive decisions rendered by evaluators in each instrument category.

CVSA and Polygraph Data Evaluations: The CVSA chart patterns were created by attaching the DAT recorder/player to the Aphex Model 120A distribution amplifier and then connecting the amplifier's output, at 65 ohms, to the CVSA instrument's microphone input jack. The use of the distribution amplifier was necessitated because of an impedance matching problem. Unlike most devices with a microphone input, and an input impedance of several thousand ohms, the CVSA's microphone input impedance is only about 600 ohms. If a device with a larger output impedance, such as the TASCAM DAT with 1000 ohms output impedance, or even a standard cassette tape recorder/player like those commonly used by law enforcement officers, is connected directly to the CVSA with a lower input impedance, the microphone line input amplifier is driven into oscillation distorting any signal input. Any amplifier stages following the input stage then only exacerbate the distortion. When the distorted signal finally arrives at the stage powering the chart stylus, which actually prints the voice pattern on the chart paper, the voice pattern that is printed may have no correlation at all with the original signal.

The polygraph charts were created by going to each of the 77 subject's polygraph folders, kept in a computerized database, and printing the charts using a Hewlett Packard LaserJet 4Plus printer.

.....

Results

The CVSA has been used for several years in the field by practitioners who have developed practices and standards which control the conduct of a VSA (Voice Stress Analysis) examination, which data is evaluated, and the evaluation practices employed in formulating a decision regarding an examinee's status, that is, truthful or deceptive. The goal of this study, was in part, to evaluate the accuracy of the CVSA in this context. A secondary goal was to then evaluate VSA data collected using nonstandard procedures, that is procedures that would not be sanctioned by field practitioners or the NITV, the organization responsible for conducting training using the CVSA. The purpose of analyzing the other data was to evaluate VSA limits using the CVSA and to see if accuracy and/or error rates could be improved using alternative procedures.

For the purposes of this study however, only the chart two CVSA results are compared to the Vericator examiner decisions and to the polygraph examiner decisions presented below. Other results and analyses will be reported later in a more extensive review of this studies findings. Figure 1 shows the accuracy of the CVSA, Vericator, and polygraph examiner decisions. Analyses found that both the polygraph and the Vericator examiner opinions discriminated between truthful and

Truthful/Deceptive Opinions Only With No Inconclusives Allowed

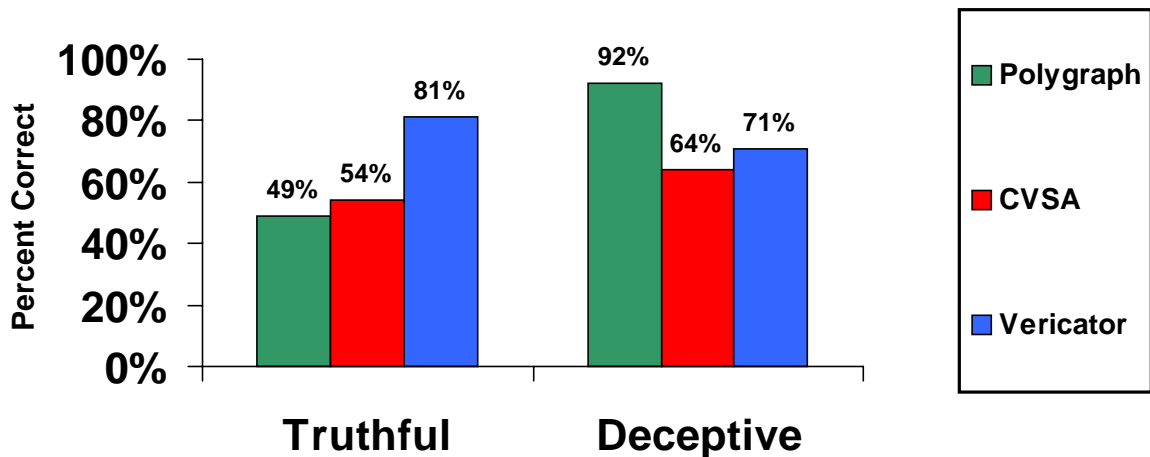


Figure 1

deceptive subjects at a high statistically significant rate, for polygraph, $\chi^2(1, N = 69) = 13.911, p < .000$, and for the Vericator, $\chi^2(1, N = 76) = 20.715, p < .000$. Concurrently, the CVSA results were not statistically significant, $p > .169$.

.....

The NITV teaches its examiners that inconclusive decisions are not acceptable. However, logic would dictate that an answer might not always be at hand. Therefore, the examiners were instructed that after making a truthful or deceptive opinion, they were to once more review the data and if warranted change their opinion from either truthful or deceptive to inconclusive. Figure 2 illustrates the distribution of the three evaluator's decisions when the inconclusive decisions were classified as non-decisions and excluded.

Again, both the polygraph and Vericator examiners were able to accurately discriminate between truthful and deceptive subjects at a rate exceeding chance. The polygraph examiner achieved an overall accuracy rate of 74%, $\chi^2(1, N = 58) = 12.796, p > .000$. Concurrently, the Vericator examiner's overall accuracy in this case was 84%, $\chi^2(1, N = 44) = 19.684, p > .000$. Once more, the CVSA examiner failed to achieve an

Truthful/Deceptive Opinions With Inconclusive Decisions Excluded

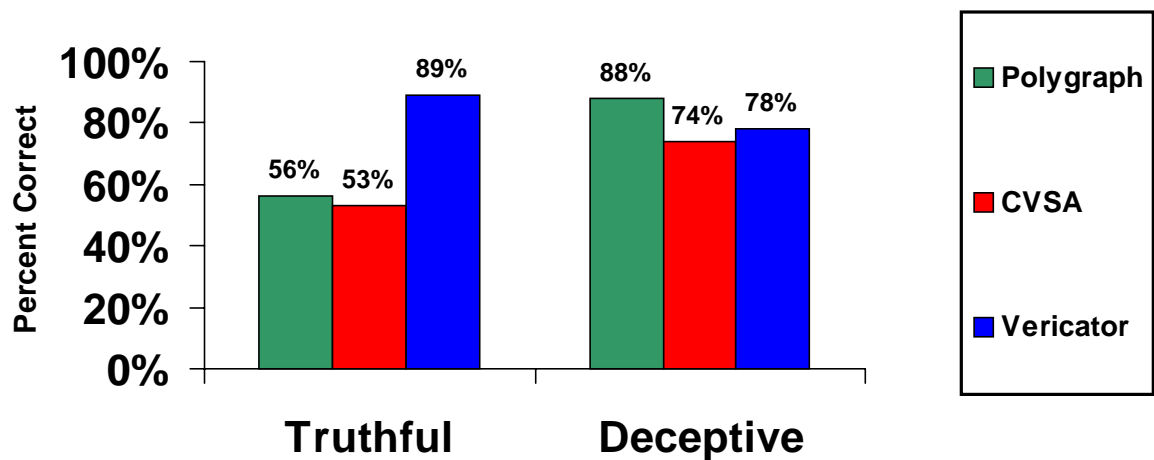


Figure 2

overall rate of accuracy exceeding chance, 64%, $\chi^2(1, N = 42) = 3.109, p = .078$.

Discussion

The preliminary results above strongly suggest that compared to the polygraph and the CVSA, the Vericator technology may have some distinct advantages. For instance, using the Vericator technology the examiner realized a higher rate of accuracy for the classification of both truthful and deceptive subjects compared to the results achieved by the CVSA examiner. Although it appears that the polygraph

•
•
•
•
•
•
•
•

may be superior for the classification of deceptive subjects, it is somewhat misleading as the higher rate of accuracy is achieved only by misclassifying a greater number of truthful (i.e., innocent) subjects also as deceptive. False positive errors may be acceptable in some settings, for instance in a security context where even one foreign agent could do grave damage. However, the vast majority of users would probably be more willing to accept a false negative error that is, allowing a few guilty people to escape rather than falsely accusing one innocent person.

At least within the limits of this study, the Vericator software provides a more even balance between accurately classifying innocent and deceptive subjects in a criminal Justice context. Minimally, these results call for increased research to explore the limits and contributions to be made by this interesting new technology. The Vericator software is less invasive compared to the polygraph and much more easy to interpret compared to the data created by a CVSA instrument.

The poor showing for the CVSA instrument was surprising as the examination format; the conduct of the examinations, and the use of real world issues should all benefit that procedure. If on the other hand, the CVSA is not as the NITV claims, an improvement on the older PSE technology, then the results seen here are very similar to those achieved by other researchers who examined the PSE more than fifteen to twenty years ago.

Certainly if nothing else these findings are provocative and should spark increased debate among practitioners, and hopefully motivate researchers interested in credibility assessment to examine this technology further.

Bibliography

Barland, G. H. (1988). The polygraph test in the USA and elsewhere. In A. Gale (Ed.), The Polygraph Test: Lies, Truth and Science (pp. 73-95). London: Sage.

Horvath, F. (1988). The utility of control questions and the effects of two control question types in field polygraph techniques. Journal of Police Science and Administration, 16(1), 1-17.

Palmatier, J. 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. J. (1999, June) The CVSA: Modern Technological Innovation or "The Emperor's New Clothes"? GP Solo and Small Firm Technical Update. American Bar Association: Chicago, Ill.

Podlesny, J. A. & Raskin, D. C. (1978). Effectiveness of techniques and physiological measures in the detection of deception. Psychophysiology, 15(4), 344-359.

Reid, J. & Inbau, F. (1977). Truth and deception: The polygraph (lie-detector) technique. Baltimore: Williams & Wilkins.