

Frank Horvath,<sup>1</sup> Ph.D.

## Detecting Deception: The Promise and the Reality of Voice Stress Analysis

---

**REFERENCE:** Horvath, F., "Detecting Deception: The Promise and the Reality of Voice Stress Analysis," *Journal of Forensic Sciences*, JFSCA, Vol. 27, No. 2, April 1982, pp. 340-351.

**ABSTRACT:** Within the past decade a number of so-called voice stress analyzers have been marketed for law enforcement and forensic science purposes. These devices are said to extract from the vocal spectrum a subaudible microtremor signal that is useful in detecting stress in a speaker's voice; thus, it is claimed these devices have great utility as lie detectors and are as accurate as the traditional polygraph instrument. A review of the evidence now accumulated about these devices shows that the evidence for the existence of a microtremor in the voice is problematic and that the capability of these devices in detecting stress is equally questionable. Without exception, however, the scientific evidence reported to date shows that voice stress analyzers are not effective in detecting deception; none of these devices has yet been shown to yield detection rates above chance levels in controlled situations. A brief comparison of voice stress analysis and polygraphic testing as methods of lie detection is made.

**KEYWORDS:** criminalistics, lie detection, voice analysis

In the lie detection field, the most widely publicized development in the past decade has been the so-called voice stress analyzer. In advertisements in popular magazines and in various trade and professional journals voice stress analyzers have been marketed as "truth machines"—devices capable of detecting lies with an accuracy that equals or exceeds that of the more traditional polygraph.

There are now some four or five different voice stress analyzers on the market. The prototypical instrument, and the one most prominently advertised, is the Psychological Stress Evaluator (PSE). The PSE was first marketed in 1971 by two former military intelligence officers who reportedly developed the device for the purpose of carrying out "lie detection" tests in a covert manner, or at least in a manner that did not require attached sensors. According to its manufacturer, Dektor, Inc., the PSE detects and measures subaudible and involuntary frequency modulations (FM) that are superimposed on audible voice frequencies. The frequency modulations, whose strength and pattern are inversely related to the degree of stress in a speaker at the moment of utterance, are said to result from minute oscillations of the muscles of the voice mechanism. Such oscillations, known as physiological tremors [1], are believed to be under control of the central nervous system during nonstressful periods. As stress is imposed, however, the autonomic nervous system gains dominance, resulting in a suppression of the microtremors. This suppression, indicative of emotional stress, is displayed by the PSE as a characteristic blocked or rectangular wave form.

The PSE processes voice frequencies preserved on a normal tape recording, using elec-

The opinions or assertions contained herein are the private views of the author and are not to be construed as official or as reflecting the views of the American Academy of Forensic Sciences or the *Journal of Forensic Sciences*. Received for publication 9 Oct. 1981; accepted for publication 23 Oct. 1981.

<sup>1</sup>Associate professor, School of Criminal Justice, Michigan State University, E. Lansing, MI 48824.

tronic filtering and frequency discrimination techniques. The stress-related FM patterns, displayed on a moving strip of heat-sensitive paper, can be processed in four different modes of display for either gross or detailed analysis; because the recovery of the FM indicator spontaneously occurs with the removal of the stressing stimulus, stress in either narrative or monosyllabic speech can be evaluated [2].

Since the development of the PSE a number of other voice stress analyzers have appeared on the market. According to the advertising literature about these devices they also detect a subaudible microtremor in the voice; thus, the theoretical physiological basis for these devices is identical to that claimed for the PSE. Some of them, however, are engineered so as to obviate the need for a graph-recorded display. Instead, they produce a direct, instantaneous analysis of the voice microtremor and signal "stress" by means of a series of flashing lights or a digital readout [3-5].

Although voice stress analyzers have other obvious applications, they are primarily marketed as a technological breakthrough in the field of lie detection. Because contactual sensors are not necessary and because a subject need be neither present nor even aware that he or she is undergoing a lie detector test, the voice stress devices are reported to be more versatile than, yet as effective as, the traditional polygraph. In fact, it is the purported versatility of those devices and their apparent usefulness in noncontemporary and covert situations that have captured the imagination of the popular media; for example, it has been reported by proponents of voice stress analysis that Lee Oswald was truthful in his denial of shooting President Kennedy [6], that President Carter lied about Bert Lance, and that Ted Kennedy told the truth about Chappaquidick [7].

The purpose of this paper is to discuss and analyze the major empirical evidence pertaining to the claims made about voice stress analysis, in particular, the assertion that voice stress devices are effective in lie detection. Because voice stress devices are usually compared to the polygraph in the research literature (as well as in advertising literature for the voice devices) a limited comparison of results obtained with those two instruments will be made. Before discussing that research, however, it will be useful to discuss briefly some of the other claims made about voice stress analyzers.

### **The Microtremor Theory**

Voice stress analyzers are said to detect subaudible, low-frequency modulations in the 8 to 12 Hz range in the voice. There have been several acceptable attempts to test that claim. Shipp and McGlone [8] found no electromyographic evidence of low-frequency tremors in the laryngeal muscles in the vocalization of either truthful or deceptive utterances. Similarly, McGlone and Hollien [9] spectrographically analyzed speech samples of subjects who read a passage in an unstressed condition and of subjects who read a passage while receiving a series of electrical shocks; they found no low-frequency energy in the speech samples of either group of subjects. Inbar and Eden [10], however, have reported that their research, in which electromyographic recordings were correlated with frequency changes in the voice spectrum, does suggest the existence of low-frequency voice tremors generated by the central nervous system. Thus, the evidence supporting the premise on which the voice stress analyzers are based is not well developed and is certainly not compelling. Nonetheless, even if the microtremor explanation is incorrect, that would not necessarily imply that the devices do not detect some vocal manifestation related to emotional stress.

### **Detection of Stress**

There have been a number of studies carried out to determine the relationship between what the voice stress analyzers detect and accepted traditional indicators of emotional stress. Many of these studies were well-controlled, reliable assessments; the results, however, have

been mixed. VanDercar et al [11], for instance, reported that they were unable to replicate their own findings of a relationship between PSE voice stress measures and heart rate and A-State scores from the State Trait Anxiety Inventory. Similarly, Brenner et al [12] were unable to obtain consistent results with a PSE voice stress analyzer in detecting stress caused by deception and that caused by performance of mental arithmetic tasks; the latter was related to voice stress patterns whereas the former was not. Lynch and Henry [13] found that PSE voice stress patterns were not effective in the identification of either stressful or un-stressful words spoken by 43 college students. On the other hand, Borgen and Goodman [14] found systematic changes in PSE voice stress measures with the Stroop color/word conflict task; those changes appeared to accompany changes in other psychophysiological measures. Other investigators have also reported a relationship between voice stress measures and indicators of stress, particularly when self-reports of subjects are the criteria [12, 14-17]. Thus, the available literature does not demonstrate that voice stress analyzers clearly and unfailingly detect emotional stress; the research results have been very inconsistent and the issue needs much more research before it will be settled. It is possible, furthermore, that such research may show the voice microtremor to be a voluntarily controlled component of the voice that is related to stress and anxiety in a largely unpredictable way; the reports of Inbar and Eden [10], VanDercar et al [11], and Brenner et al [12] suggest such an outcome.

#### **Detection of Deception: Controlled Studies**

Unlike the research reported pertaining to other claims made about voice stress analyzers, the well-controlled studies in which lie detection has been at issue have yielded consistent results: none of them has shown that the devices are effective in detecting deception. Because there are relatively few of these studies, they will all be discussed here briefly.

The first scientifically acceptable study of the validity of voice stress devices in lie detection was reported by Kubis [18] at Fordham University in 1973, about three years after the prototypical instrument was marketed as a lie detector. Kubis designed an elaborate series of studies to determine the relative effectiveness of the polygraph, the PSE, and another voice stress device, the Voice Stress Analyzer (VSA), produced by Decision Control, Inc., in detection of deception. Kubis's study consisted of a "mock crime paradigm" in which some college students were assigned the role of thief, some were the lookout, and some the innocent bystander. Kubis's findings showed that neither the PSE nor the VSA was effective in discriminating between the three student roles. The PSE yielded an accuracy of 32% (27/85) in detecting individual's roles in one portion of that study and 38% (24/63) in detecting roles within each three-student grouping in another portion, against chance expectancy of 33% in each case; the VSA showed an average accuracy of 36% (39/108) in those same situations. On the other hand, polygraphic analysis in Kubis's experiment showed a highly significant overall detection rate of 76%. It is of some interest to note here that Kubis also found that the conditions of his study were sufficiently motivating to produce observable behavioral differences between truthful and deceptive subjects; persons who evaluated only the subjects' behavior during testing were able to discriminate between truthful and untruthful subjects with greater accuracy (53%) than was obtained with the PSE or the VSA.

In another study, Barland [19] carried out two small-scale projects to determine the accuracy of the PSE in lie detection. In the first, he had a group of 16 college students conceal information; they were then tested with the PSE to determine if the concealed information could be detected. The results of that experiment showed that the accuracy of the PSE was at chance levels, 6.25% (1/16), a finding that Barland believed to be related to the students' lack of motivation to deceive. To investigate that hypothesis, Barland, in his second project, tested 14 actual criminal suspects—believed to be highly motivated to deceive—with the PSE and the polygraph. He reported initially that the PSE appeared to indicate reliable changes in the voice associated with deception and that the PSE was more effective in conditions of

heightened motivation. In another study, larger in scale and more carefully executed, however, Barland [20] found that the accuracy of the PSE (averaging 51%) did not exceed chance levels (0.50) in detecting deception in criminal suspects, whereas in the same circumstances the polygraph yielded an accuracy of about 90%. Thus, Barland's original hypothesis about the effect of motivation on the effectiveness of voice stress analysis was not supported in his own research.

Nachshon and Feldman [21] reported a series of studies designed to investigate the effectiveness of voice stress analysis in detecting concealed information. In one portion of their study, 20 college students concealed cards chosen from a deck of six cards. The students were then tested with the PSE; evaluation of the PSE data by three trained evaluators yielded an average accuracy rate of 30%, a result not significantly greater than chance expectancy. In another portion of their study, Nachshon and Feldman evaluated the accuracy of the PSE in detecting cards concealed by 19 criminal suspects who were undergoing polygraph examinations. In those presumably more motivating circumstances, Nachshon and Feldman found that the PSE yielded an average accuracy of 19%, ranging between 15% and 26% for the three evaluators; the PSE did not produce an accuracy greater than chance expectancy (0.20).

Two other laboratory-based studies of the accuracy of voice stress analysis were reported by Horvath [22,23] at Michigan State University. In the first study, 60 college students, 30 male and 30 female, attempted to conceal numbered cards chosen from a deck of five cards while undergoing simultaneous PSE and polygraph testing. Analysis of PSE response data and polygraphic response data, the galvanic skin response (GSR) in particular, was carried out by two trained evaluators. The detection rates obtained with the PSE averaged 22.5% against chance expectancy of 0.20 and were not significantly affected by subjects' sex, repeated trials of testing, simultaneous use of polygraphic and voice stress equipment, or differences between the two trained evaluators of the PSE data. In that same study, detection rates obtained in scoring GSR responses averaged 68.6% (in the first trial of testing only) against chance expectancy of 0.20, and in all cases the rates were significantly greater than chance.

Horvath [23] also investigated whether or not the accuracy of the PSE could be enhanced by increasing the subjects' motivation to deceive. In this study 64 college students were promised a reward for successfully completing a task involving the concealment of a numbered card chosen from a deck. In spite of the evidence showing that the subjects were indeed considerably motivated by the reward, that motivation did not increase detection rates obtained with voice stress analysis beyond chance levels; the PSE averaged only 18% correct detections against chance expectancy of 0.20. On the other hand, detection rates obtained with only the GSR in that same study averaged 52%, significantly exceeding chance levels.

It is of some interest to note that in both of the studies reported by Horvath, voice stress analysis yielded lower detection rates than were obtained by analysis of each of the three physiological measures recorded polygraphically—GSR, respiration, and cardiovascular activity [24]. Thus, Horvath's findings were remarkably consistent with those reported by Kubis [18]; when evaluated in similar contexts voice stress analysis did not yield an accuracy similar to that obtained with the polygraph.

In a recently reported study, Brenner et al [12] carried out a lie detection task in which the PSE was used to detect ten items of personal information concealed by 20 college students. The students were offered a reward if they were successful in avoiding detection of the items. By random scoring of the subject's PSE responses, an average of 20% of the concealed items would have been detected. The results of the analysis showed that the actual detection rates were not significantly different from chance levels. Depending on the manner in which the PSE responses were scored the detection rates varied between 18.6 and 21.0%. When only the clearest voice stress charts were separately evaluated detection rates remained at chance

levels; in spite of the large variation noted in the nature of the stress responses, the variation was not related to the experimental manipulations. Brenner et al point out, moreover, that when used to detect concealed information in the same manner as they used the PSE, the polygraph has yielded detection rates as high as 100%.

### **Objections to the Controlled Studies**

The studies discussed to this point represent the bulk of the reliable evidence reported to date about the effectiveness of voice stress analyzers in detecting deception. Although that evidence clearly does not support the claims made about voice stress analyzers, the proponents of such devices challenge that evidence on two major grounds. First and perhaps foremost among the proponents' arguments is that most of the reliable evidence has been laboratory-based and has involved mere "game playing" situations with low levels of jeopardy. Since, they say, the devices were not designed to be used in such situations it is not surprising that they would be found to be ineffective in them. Although this argument has some ostensible merit, there are a number of points made by the proponents themselves that mitigate its authority.

According to the manufacturers and proponents of the voice stress devices they have the capability to detect absolute stress levels [3-5,25]. Presumably, such a claim suggests that not only can the devices detect whether stress is present but also the degree of stress, a claim, by the way, which is a significant feature of the training program of the manufacturers.<sup>2</sup> If, of course, this claim were true, then whether or not the testing situation involved low or high levels of stress would generally be of little consequence; that is, if the devices did detect absolute stress levels one would expect to be able to determine easily, for instance, which of a group of items yielded the greatest degree of stress. The evidence does not suggest that possibility.

On the other hand, if there is a certain degree of jeopardy (stress) necessary to obtain valid results with the voice stress devices, as the proponents also claim, what is the threshold and what is the criterion by which one determines it? Is it always present in real life and never in laboratory situations? Those issues have not yet been addressed by the proponents, nor is there any information given about them in training manuals and other material offered by the proponents [25].<sup>2</sup>

In explaining how the prototypical voice stress device was developed, its coinventor has stated that "We set up a known stress/nonstress situation on tape and ran experimental charts with various types of signal processing to attempt to detect any change which may occur which was notable in the stress area which would differ from the representation in the unstressed area and, proceeding with this into refinement, we were able to increase the effectiveness of this by changes in signal processing [26, p. 111]." That testimony about how the first voice stress device was developed and perfected appears to be at odds with the proponents' claim that the device was not designed to detect stress in experimental situations. In that same testimony, in response to a question about what validation studies were done, it was further stated that there was "extensive use of the 'To Tell the Truth' program as broadcast over television, simply because it provided us with a difficult situation where jeopardy of the usual type of lie detection jeopardy was not present. We had singular success with this [26, p. 112]. . . . I think the 'To Tell the Truth' [accuracy] was something like 94.7 percent [26, p. 145]." At a later point in the testimony addressing the validity of the device in situations where there is less than real-life jeopardy, it was stated that "the PSE doesn't do particularly well in this unless the individual is specifically trained for that application. Our salesmen can do it. The usual PSE examiner is not taught to do that. That is not what they're using it for [26, p. 129]."

<sup>2</sup>Personal notes from PSE training course, sponsored by Dektor, Inc., Springfield, VA, December 1975. See also Refs 2 and 4.

The inconsistency between the claims for and about voice stress devices, and the proponents' major objection to the laboratory-based studies, is obvious. On the one hand, the devices were not designed to be used in experimental situations; on the other, that is precisely how they were developed and validated. On the one hand, the devices are not effective in experimental situations because the stress levels are too low; on the other hand, it is not the devices that are at fault here, since salesmen can apparently be taught how to detect low-jeopardy lies. Thus, it is far from clear why those who have been trained to actually use voice stress devices in detecting deception have been unable to demonstrate their validity in controlled situations.

A second objection made to the studies about voice stress devices is that the operators in those studies did not use valid chart reading techniques, that is, that they did not analyze the response data in a proper manner [25]. This objection, like the one already discussed, does not square with the evidence. In each of the lie detection studies discussed previously [12,18–24] the response data were analyzed by evaluators trained and certified by a major manufacturer as being qualified to interpret data. Moreover, it is clearly expressed in those studies that the criteria advocated by the manufacturer were indeed those that were applied in analyzing the data.

A more telling point in response to this objection, however, is that regardless of how data are scored there is very little agreement among raters on voice stress responses. Correlation coefficients in the reported studies are generally quite small and strongly suggest, as Brenner et al have reported, that “unreliability is built into the basic scoring procedure and [is] not simply a reflection of inexperience [12, p. 352].” In fact, there are at least two reports that show that untrained or inexperienced evaluators agree as often as, if not more often than, experienced evaluators, although neither judge response data very accurately. In one of these reports, an unpublished study by Worth and Lewis,<sup>3</sup> it was found that an untrained evaluator had higher detection rates in a laboratory situation than a trained evaluator, 58% versus 50% where chance expectancy was 0.25. In the second, more recent report it was found that a manufacturer's employee who trains voice stress operators did considerably worse (less agreement with a polygraph-based criterion) in analyzing response data from real life situations than did two trained but less experienced evaluators. In no case did the three raters agree in even 50% of their evaluations, the agreement rate between the two inexperienced evaluators being only 32% whereas the agreement rate between the employee and the other two evaluators averaged only 40% [27]. These low rates of agreement, of course, merely reinforce the various findings showing the low validity obtained with the voice stress devices.

### Controlled Field Studies

Two reliable, independent studies deserve special mention at this point. Each of these studies involved an evaluation of the PSE in field situations (criminal testing); therefore, there can be no objection to them on the ground that they were carried out in an artificial setting. Furthermore, in one of these studies the response data were evaluated by three different persons, all of whom were certified as competent analysts by the manufacturer. Thus, there is little doubt that the data were evaluated in a manner consistent with the manufacturer's guidelines.

In the first study, reported by Barland [20], 66 criminal suspects were tested using both polygraphic and PSE equipment. There was no significant relationship between the scores derived from analysis of polygraphic data and those derived from analysis of PSE data. More important, Barland assessed the accuracy of his PSE-based decisions by three criteria: con-

<sup>3</sup>J. Worth and B. Lewis, “An Early Validation Study with the Psychological Stress Evaluator (PSE),” unpublished paper, Washington and Lee University, Lexington, VA, 1972.

fessions or guilty pleas in court, decisions made by a panel of legal experts on the basis of written documentation in each case, and the outcome in each case in which there was an independent judicial decision made. Barland's results showed that the accuracy of the PSE was not significantly greater than chance expectancy (0.50). Regardless of which of the three criteria was used as the standard of ground truth, the accuracy in each instance averaged about 50%.

A more recent field-based evaluation of the PSE was carried out by the Department of Commerce in Virginia. In that study, the Department of Commerce, the Virginia State Police, and Dektor, Inc., agreed on the design of a study in which blind PSE evaluations were compared to results obtained in polygraph examinations of persons involved in actual criminal investigations. For that purpose Dektor, Inc., trained and certified two operators. Those two operators and an employee of Dektor independently analyzed PSE data in 40 cases in which complete data were available. When the PSE results in those 40 cases were compared to the polygraph-based outcomes there was no significant association between the conclusions reached by the two methods; the PSE results agreed with the polygraph outcomes, on the average, in 39% of the cases, compared to the 33% agreement that would be expected by chance [27, p. 16]. It is of some interest to note also that the PSE operators performed slightly better when their results included PSE data that they claimed were "unusable" than when those data were excluded and that "substantially the worst performance was recorded by the Dektor employee" [27, p. 17]. Thus, the authors of this report conclude that "by all conventional standards of proof we have to regard the validity and reliability of the Psychological Stress Evaluator as unproven. Indeed, it appears that by and large its validity and reliability are not only unproven, but rather are disproven [27, p. 19]."

#### **Analysis of the Reports of Voice Stress Users**

Although all of the reliable, independent studies have shown consistent results—whether they were laboratory- or field-based—there are several other reports that, according to the proponents of the voice stress devices, support their claims for the effectiveness of those devices in detecting deception. None of these reports, however, meets generally accepted standards within the scientific community; for that reason alone they are of dubious value. Nonetheless, because these reports are the only ones that buttress the proponents' case they will be briefly discussed here.

In 1972, M. Kradz, in an unpublished paper,<sup>4</sup> reported that he had carried out both polygraph and PSE testing simultaneously on 42 criminal suspects; one additional suspect was tested with the PSE only. Of the 43 suspects tested, 27 were said to be "cleared of suspicion" on the basis of the PSE testing; 21 of those were corroborated as innocent by "independent investigation." Of the 16 suspects "not cleared" by the PSE the guilt of each was said to be established by additional investigation or confession or both. Kradz claimed that his results showed that "100% accuracy was produced in those 36 subject examinations for which complete and concrete corroboration was, or later became, available."

Unfortunately, Kradz's report did not reveal a number of details about his method that are critical to a determination of what his findings might actually suggest. For instance, it was not indicated precisely how the actual guilt or innocence (ground truth) was established for each of the suspects, nor was it clear who carried out the "independent investigation" that apparently established the ground truth criterion Kradz used. When asked how he had ascertained ground truth in his study, Kradz testified that he used "independent physical evidence" such as "fingerprints, finding of the weapon, the deceased, stolen property, and questioned documents [26, p. 196]." And, when asked if he had used confessions to

<sup>4</sup>M. Kradz, "Psychological Stress Evaluator: A Study," first version of an unpublished paper distributed by Dektor, Inc., Springfield, VA, dated 1972.

establish ground truth, Kradz replied: "Oh, no, not even eyewitnesses [26, p. 197]," although his written report states that an "admission of guilt" was used to corroborate guilt in 13 of 16 cases in which suspects were "not cleared." In fact, according to the written report, in 25% of those cases an "admission of guilt" was made before the "independent" investigation. Kradz further testified that he did not use the outcome of trials in which the suspects were involved because "in two cases we disagreed with that [26, p. 197]." He said he himself determined when the evidence was sufficient to establish that the "PSE was worthy of use in criminal justice [26, p. 200]." The latter statement suggests that the independence of Kradz's "independent investigations" is questionable.

Although Kradz has not yet clarified the details of his method,<sup>5</sup> another version of his report,<sup>6</sup> which is distributed as the original study "reproduced verbatim in its entirety," further confounds the issues. This report describes a method and a number of critical details that are different from what the original report described. The second version, for instance, reports that an unspecified number of the PSE charts were evaluated "in the blind," whereas the first version pointed out that both the subject and the examiner (Kradz) discussed during the testing what was indicated on the PSE charts. Moreover, the second version is even less clear about how ground truth was established than was the first version.

Because of the conflicts between Kradz's original report, what he has stated in testimony, and the recently distributed copy of his report, it is not possible to determine what Kradz actually did. The serious and unexplained methodological deficiencies in the Kradz study clearly indicate that that study does not meet generally accepted scientific standards; his reported findings, therefore, are of questionable value in assessing the validity of voice stress analysis.

One of the coinventors of the PSE has reportedly claimed that the device is "96.78 percent effective [7, p. F-2]." That claim apparently is based on a statistic reported by another voice stress proponent, Heisse, as a result of a study he carried out to investigate the "reliability and validity" of the PSE.<sup>7</sup> In his study, Heisse selected 53 cases (contributed by PSE users) in which the PSE was used to determine the truthfulness of the suspects (some of the "suspects" were applicants for employment, not persons involved in criminal investigations). Twenty-six of the suspects were apparently known to have been deceptive (to have shown "some form of deception") during their PSE testing; 27 were known to have been truthful. In each case ground truth apparently was established by a confession that indicated either the deception of the guilty suspect or the truthfulness of the innocent suspect. Of the 53 separate suspects tested, 25 of them were involved in three separate investigations.

Heisse asked 12 PSE users to evaluate the PSE charts of each of the 53 suspects and to determine whether each suspect was "truthful" or "deceptive." He reported his findings thusly: "There are 258 acceptable interevaluator replies. Among the replies there are 10 errors. . . . Hence, the interevaluator reliability is 96.12 percent." Later in his paper he reports: "The compliance between evaluators and the known results with 258 evaluation replies is 96.12 percent. If examiners are included in this group . . . the reliability jumped to 96.78 percent."

Heisse's report, like that of Kradz, fails to disclose a number of important methodological details. Precisely how the cases were sampled, for example, is not revealed, nor is any procedure identified that would have ensured the independence of those persons who evaluated the response data. Since 25 of the suspects were involved in the same three investigations, it

<sup>5</sup>Personal communication with M. Kradz, Dektor, Inc., Springfield, VA, 24 Jan. 1980, 20 Feb. 1980, and 10 March 1980.

<sup>6</sup>M. Kradz, "Psychological Stress Evaluator: A Study," second version of an unpublished paper distributed by Dektor, Inc., Springfield, VA, dated 1971.

<sup>7</sup>J. Heisse, "Audio Stress Analysis: A Validation and Reliability Study of the Psychological Stress Evaluator (PSE)," unpublished manuscript dated 1 Feb. 1976, available from the author, 144 Cliff St., Burlington, VT.



is certain that the tests carried out on those persons in each investigation were not independent; yet, Heisse does not indicate how that issue was dealt with, if at all. Although there are other serious methodological problems evident in the Heisse study, it is also the case that his findings were not correctly interpreted. Heisse himself, for instance, has stated that contrary to what his report suggests his findings only deal with the issue of reliability—how consistently his evaluators interpreted his data—and not with validity.<sup>8</sup> But, proponents often use his statistics to support their claim that voice stress analysis is 96.87% accurate. Judging from what was reported by Heisse and Kradz, such a claim is unfounded.

A final study claimed to support voice stress analysis is a paper reported by Dahm,<sup>9</sup> who sent questionnaires to 423 users of the PSE; of those, 46 responded to questions about several characteristics of their use of the PSE. Dahm's major findings were reported as follows. First, he said that polygraph and PSE examinations were in agreement 5037 times in 5045 cases, "for a correlation of 99.84%." Second, "Based upon 10,202 PSE examinations . . . there was not one case in which the PSE had been found in error [28, p. 231]." It is, of course, clear that Dahm's data represent merely the unsubstantiated opinions of only a small number of PSE users; they are not sufficient to indicate whether or not voice stress analysis is a valid means of detecting deception.

Thus, the Kradz, Heisse, and Dahm reports constitute at best merely testimonial, not scientific, evidence of the effectiveness of voice stress analysis. The merits of those studies notwithstanding, however, it is interesting that all of them were reported by proponents of voice stress analysis after 1971; neither the manufacturers nor the other proponents of voice stress devices have yet produced a report of research which was carried out *before* the devices were publicly marketed. The developmental research supporting the validity of the devices in lie detection is, curiously, not available. It is also important to point out that the findings in the proponents' studies regarding the accuracy of voice stress analysis have not yet been replicated in any objective, independent research. One manufacturer, asked for proof of the validity of his voice stress analyzer, reportedly sent to author B. Rice [29] a packet of ten studies, all of them very favorable. The studies were unpublished; two were apparently performed by an independent testing firm. When Rice investigated the firm he reportedly found that its president was the manufacturer of the voice analyzer. When that manufacturer was asked who did the other studies, "He replied cheerfully, 'I did. I did them all' [29, p. 72]."

### Discussion: The Polygraph and Voice Stress Analysis

In the formative years of field lie detection, a number of the proponents of the method claimed great success using not a polygraph but merely a measure of one physiological response system. As examples, Marston [30] advocated the use of a "systolic blood pressure test," Benussi [31] a test based on respiratory patterns, and Summers [31] a test based on a measure of electrodermal response (GSR). Although it has been demonstrated today that each of those response systems is useful in detecting deception [32,33], it has also been shown that each makes a separate and independent contribution to the process of lie detection [32,33]. Thus, the polygraph, which simultaneously monitors a number of physiological systems, represents a technological advance over the devices used earlier. It is also certain, however, that as important as it is to record a number of response systems, the manner in which polygraph testing is administered and the way in which polygraphic data are interpreted are of at least equal importance [34,35]. It is recognized today that lie detection is a difficult, complex, and subtle process in which the polygraph instrument itself merely provides the foundation for the structure of what is called the polygraph technique, the art, if you will, of detecting deception with a polygraph instrument. Hence, since the polygraph

<sup>8</sup>J. Heisse, Burlington, VT, personal communication, 11 March 1980.

<sup>9</sup>A. Dahm, "Study of the Field Use of the Psychological Stress Evaluator," unpublished paper distributed by Dektor, Inc., Springfield, VA., undated.

itself does not detect lies, the technique is not infallible. There is, not surprisingly, considerable controversy about how valid (accurate) the technique is; nonetheless, even the most severe critics acknowledge that the evidence clearly shows an accuracy sufficient to justify the use of polygraph testing for certain purposes [36].

Voice stress analysis, according to its proponents, promises a technologically advanced, simple, easy, almost infallible method of detecting deception, that uses, moreover, information collected from only one response system, the voice [25]. Thus, some of the claims made about voice stress devices are not entirely dissimilar to those made in the formative years of polygraphic instrumentation. There, however, the similarity ends. There is no compelling evidence that any voice stress device actually detects a signal (physiological change) that is clearly and dependably related to stress resulting from deception or any other cause. In fact, the reliable evidence that does exist shows that there is no relationship between what the voice stress devices detect and deception-induced stress. Given those facts, by the way, federal and state courts and state regulatory agencies [27,37,38] have recently ruled against proponents of the voice stress devices who have sought the same recognition afforded those who use polygraphic instrumentation.

The developmental history of the polygraph technique shows a conscious, continuing concern with standards of selection and training of polygraph examiners, in clear recognition of the fact that the technique is a complex endeavor in which the polygraph instrument plays a necessary but relatively subordinate role to the technique itself [29-32]. Voice stress proponents deny that developmental history and maintain that voice stress analyzers represent advanced technology that, among other things, "simplifies chart-reading and greatly reduces both the training time required and the subjectivity of the chart reading [25, p. 64]." The evidence, of course, most clearly does not support such assertions. It is important to point out, however, that even if the evidence showed a dependable relationship between deception and what is recorded by voice stress analyzers, the historical, scientific, and practical lessons and developments in the lie detection field are proof enough of the falsity of such assertions as those made by proponents of voice stress devices. In other words, there can be no device, no instrument, no new technology that makes lie detection any less complex than it has already been shown to be.

In summary, the promise of voice stress analysis in the lie detection field is not and may never be a reality. All of the reliable evidence now available shows that none of the voice stress devices is useful in detecting deception; the fact that the precise relationship between the components of the voice spectrum and emotional states has not been adequately specified suggests a formidable obstacle to be overcome before analysis of the voice may prove of value in lie detection. The fact that voice stress devices have apparently been accepted rather uncritically by some law enforcement agencies, and for some forensic science purposes, is a development which, judging from the available evidence, cannot now be justified.

## References

- [1] Lippold, O., "Physiological Tremor," *Scientific American*, Vol. 224, No. 3, March 1971, pp. 65-73.
- [2] *PSE Orientation Course Manual*, Dektor Counterintelligence and Security, Inc., Springfield, VA, undated.
- [3] *Voice Stress Computer*, Omnitronics Research Corp., Akron, Ohio, 1980.
- [4] Bennett, R., Hardee, P., and Klobert, R., *Hagoth: Fundamentals of Voice Stress Analysis*, Hagoth Corp., Issaquah, WA, 1977.
- [5] *Voice Stress Analyser System, Mark IX-P*, Communication Control Systems, Inc., New York, 1978.
- [6] O'Toole, G., "Lee Harvey Oswald Was Innocent," *Penthouse*, Vol. 6, April 1975, pp. 45-46, 124-132.
- [7] Jenkins, J., "Voice-Analyzing Lie Detector Yields Questionable Results," *Albuquerque Journal* (newspaper), 12 Oct 1979, p. F-2.

- [8] Shipp, T. and McGlone, R., "Physiologic Correlates of Acoustic Correlates of Psychologic Stress," presented at the meeting of the Acoustical Society of America, Los Angeles, Nov. 1973.
- [9] McGlone, R. and Hollien, H., "Partial Analysis of Acoustic Signal of Stressed and Unstressed Speech," in *Proceedings of the 1976 Carnahan Conference on Crime Countermeasures*, BU 110, ORES Publications, College of Engineering, University of Kentucky, Lexington, 1976, pp. 19-21.
- [10] Inbar, G. and Eden, G., "Psychological Stress Evaluators: EMG Correlation with Voice Tremor," *Biological Cybernetics*, Vol. 24, 1976, pp. 165-167.
- [11] VanDercar, D. H., Greaner, J., Hibler, N. S., Spielberger, C. D., and Bloch, S., "A Description and Analysis of the Operation and Validity of the Psychological Stress Evaluator," *Journal of Forensic Sciences*, Vol. 25, No. 1, Jan. 1980, pp. 174-188.
- [12] Brenner, M., Branscomb, H., and Schwartz, G., "Psychological Stress Evaluator: Two Tests of a Vocal Measure," *Psychophysiology* (Madison, WI), Vol. 16, No. 4, July 1979, pp. 351-357.
- [13] Lynch, B. and Henry, D., "A Validity Study of the Psychological Stress Evaluator," *Canadian Journal of Behavioral Science*, Vol. 11, No. 1, March 1979, pp. 89-94.
- [14] Borgen, L. and Goodman, L., "Voice Print Analysis of Anxiolytic Drug Effects: Preliminary Results," *Clinical Pharmacology and Therapeutics*, Vol. 19, No. 1, March 1976, p. 104.
- [15] Wiggins, C., McCranie, M., and Bailey, P., "Assessment of Voice Stress in Children," *Journal of Nervous and Mental Disease*, Vol. 160, No. 4, Dec. 1975, pp. 402-408.
- [16] Worth, L., Lewis, B., and Raborn, G., "Presence of the Dentist: A Stress Evoking Cue?" *Virginia Dental Journal*, Vol. 52, No. 2, April 1975, pp. 23-27.
- [17] Brockway, G., "Situational Stress and Temporal Changes in Self-Report and Vocal Measurement," *Nursing Research*, Vol. 28, No. 2, March 1979, pp. 19-24.
- [18] Kubis, J., "Comparison of Voice Analysis and Polygraph as Lie Detection Procedures," report, Contract DAAD05-72-C-0217, U.S. Army Land Warfare Laboratory, Aberdeen Proving Ground, MD, Aug. 1973.
- [19] Barland, G., "Use of Voice Changes in the Detection of Deception," presented at the 86th meeting of the Acoustical Society of America, Los Angeles, Oct. 1973 (erratum sheet issued March 1974).
- [20] Barland, G., "Detection of Deception in Criminal Suspects: A Field Validation Study," Ph.D. dissertation, Dept. of Psychology, University of Utah, Salt Lake City, 1975.
- [21] Nachshon, I. and Feldman, B., "Vocal Indices of Psychological Stress Evaluator," *Journal of Police Science and Administration*, Vol. 8, No. 1, March 1980, pp. 40-53.
- [22] Horvath, F., "An Experimental Comparison of the Psychological Stress Evaluator and the Galvanic Skin Response in Detection of Deception," *Journal of Applied Psychology*, Vol. 63, No. 3, June 1978, pp. 338-344.
- [23] Horvath, F., "Effect of Different Motivational Instructions on Detection of Deception with the Psychological Stress Evaluator and the Galvanic Skin Response," *Journal of Applied Psychology*, Vol. 64, No. 3, June 1979, pp. 323-330.
- [24] Horvath, F., "Detecting Deception in the Voice: The Validity of the Psychological Stress Evaluator," presented to the American Polygraph Association, St. Louis, Aug. 1978.
- [25] Bell, A. "The PSE: A Decade of Controversy," *Security Management*, Vol. 25, No. 3, March 1981, pp. 63-73.
- [26] Testimony of A. Bell and M. Kradz, *Milano, F. v. Garrison et al.*, transcript of proceedings in the District Court of the United States for the Western District of North Carolina, Charlotte Division, C-C-79-334, Dec. 4 and 17, 1979, pp. 267.
- [27] *Report of the Department of Commerce on the Feasibility and Desirability of Licensure of Audio Stress Examiners to the Governor and the General Assembly of Virginia*, House Document No. 5, Commonwealth of Virginia, Richmond, VA., 1981.
- [28] *Polygraph Control and Civil Liberties Protection Act*, hearings before the Subcommittee on the Constitution of the Committee on the Judiciary, United States Senate, 95th Congress, 15-16 Nov. 1977 and 19 and 21 Sept. 1978, U.S. Government Printing Office, Washington, DC, 1978.
- [29] Rice, B., "The New Truth Machines," *Psychology Today*, Vol. 12, No. 1, June 1978, pp. 61-78.
- [30] Marston, W., *The Lie Detector Test*, Richard R. Smith, New York, 1938.
- [31] Trovillo, P., "A History of Lie Detection," *Journal of Criminal Law and Criminology*, Vol. 29, 1939, pp. 848-881 and Vol. 30, 1939, pp. 104-119.
- [32] Horvath, F., "Detection of Deception: A Review of Field and Laboratory Procedures and Research," *Polygraph*, Vol. 5, No. 2, June 1976, pp. 107-145.
- [33] Podlesny, J. and Raskin, D., "Physiological Measures and the Detection of Deception," *Psychological Bulletin*, Vol. 84, No. 4, Dec. 1977, pp. 782-799.
- [34] Horvath, F., "Polygraphy: Some Comments on the State of the Art," *Polygraph*, Vol. 9, No. 1, March 1980, pp. 34-41.
- [35] Reid, J. and Inbau, F., *Truth and Deception: The Polygraph ("Lie Detector") Technique*, 2nd ed., Williams and Wilkins, Baltimore, 1977.

- [36] Lykken, D., "Psychology and the Lie Detector Industry," *American Psychologist*, Vol. 29, No. 10, Oct. 1974, pp. 724-739.
- [37] Ansley, N., Ed., "PSE User Loses in Federal Court," *American Polygraph Association Newsletter*, Vol. 14, No. 1, Jan.-Feb. 1981, p. 2.
- [38] Ansley, N., Ed., "Illinois Supreme Court Upholds Polygraph Licensing Law," *American Polygraph Association Newsletter*, Vol. 13, No. 6, Nov.-Dec., 1980, p. 1.

Address requests for reprints or additional information to  
Frank Horvath, Ph.D.  
School of Criminal Justice  
Michigan State University  
E. Lansing, MI 48824