Harry Hollien,[1] Ph.D., Laura Geison,[1] M.A., and James W. Hicks, Jr.,[2] Ph.D.

Voice Stress Evaluators and Lie Detection

REFERENCE: Hollien, H., Geison, L., and Hicks, J. W., Jr., **"Voice Stress Evaluators and Lie Detection,"** *Journal of Forensic Sciences,* JFSCA, Vol. 32, No. 2, March 1987, pp. 405-418.

ABSTRACT: Proponents of vocal stress analysis systems argue that they are able to detect spoken deception by analysis of "stress" in the voice signal. Presumably, they do so by examining traces made by laryngeal microtremors which, they claim (1) exist in the voice, (2) are associated with stress, and (3) ultimately are associated with lying. However, most research that seeks to identify the relationships between microtremors and laryngeal function has produced negative results, and data on the ability of voice analyzers to detect stress from speech—or to identify spoken deception—have been negative or "mixed" in nature. Since perspectives based on available results leave a number of questions unanswered, a series of experiments has been undertaken. The first was focused on the basic acoustic/temporal correlates of stress in voice (the subject of an earlier report), the second on examination of stress by commercial voice analyzers, and the third on the detection of relatively high-risk lies by this same type of voice analysis procedure. It was found that correct stress/nonstress identifications occurred only at chance levels; the lie/nonlie identification scores were quite similar with professional "examiners" performing at about the same level of accuracy as other auditors. The following review is divided into two parts: a history of the controversy and a presentation of the two cited experiments.

KEYWORDS: forensic science, lie detection, voice analysis, stress in voice, detection of deception, voice analyzers, psychological stress evaluator (PSE), speaking fundamental frequency (SFF)

It is without question that law enforcement, security, intelligence and related agencies need an effective method for the detection of deception—one that is both valid and reliable. It is clear that the polygraph [*1,2*] alone cannot fill this need. Although useful in lie detection, polygraphy is limited by the situation in which it is employed and, especially, by the skill of the operator. Further, certain problems with its day-to-day use have been documented [*3,4*], and it should be noted that the results of polygraph tests ordinarily are not admissible in courts of law unless so stipulated. Thus, although helpful, polygraphy hardly can be considered a realistic solution to the cited problem.

Proponents of voice stress analyzers argue that their devices fill the need for a procedure that detects deception. They maintain that their techniques permit examiners to identify stress, and hence lies, solely from analysis of the acoustic signal resulting from a speech act. A number of voice analyzers now are commercially available and have been described [*4-9*]; all are purported to function as lie detectors. But are these claims valid? Can the voice ana-

Based on papers and panel discussions at the 32nd Annual Meeting of the American Academy of Forensic Sciences, 1980. Received for publication 13 Sept. 1985; revised manuscript received 25 April 1986; accepted for publication 1 May 1986.

[1]Professor and founding director, and research assistant, respectively, Institute for Advanced Study of the Communication Processes, University of Florida, Gainesville, FL.

[2]Research associate, Institute for Advanced Study of the Communication Processes, University of Florida, and staff scientist, SCI Systems, Inc., Huntsville, AL.

lyzers be used successfully to measure stress, and does stress always accompany lying? It is quite possible that Lykken [*4*] has articulated one of the key concepts relative to spoken deception. He argues that, if lies are to be detected, there must be a "lie response." That is, there has to be some measurable physiological or psychological event (stress related or not) that *always* occurs when a person lies. Until such a response has been validly and reliably identified, no one can claim to be able to detect and measure falsehoods. But has a lie response been isolated? Simple logic can be used to test this possibility as the consequences of such an occurrence would be extensive. As Lykken points out, the impact on all levels of society would be almost inconceivable—consider the effects on international relations, politics, business, the courts, and even family relationships. Is it possible that the proponents of the voice analyzers have identified the "lie response"? In a sense, they claim that they have.

Vocal Stress/Deception

At this point, it would appear useful to identify vocal stress and define its nature. First, stress often is demarcated in terms of an applied stressor. However, this definition is not very useful because, even though the stressor itself is defined, the subjects' emotional state actually is unknown. On the other hand, when emotions *are* identified, they often are simulated by actors. The reason for this investigational technique, in turn, is due to the fact that it is rarely possible to record the speech of an individual who is experiencing a particular emotion at the instant it occurs. Further, the situation is confounded because it also is difficult to obtain similar, unstressed utterances (produced by that same individual) for comparison. Hence, many of the relationships to follow are based on actors' speech or low stress states; only occasionally are data on the actual effects of stress on speech found in the literature [*10,11*]. Nevertheless, definitions of stress (and, hence, stress in voice) are possible. That is, this emotional state may be defined as a response to some sort of perceived hazard [*12,13*]. Scherer [*14*] refines the definition by suggesting that it can be either internal or external with adaptive or coping behavior required; however, Basowitz, et al. [*15*] contend that it is not "imposed" at all, but rather constitutes an individual's "response" to stressful conditions. In any case, the definition used here is one first articulated by Hicks [*16*]; that is, stress is a "psychological state which results as a response to a perceived threat and is accompanied by the specific emotions of fear, anxiety and/or anger."

But what are vocal correlates of stress states (or emotions)? It would appear desirable to review them to establish a perspective against which the claims of the voice analysis proponents can be evaluated. First, it is well accepted that it is possible to (perceptually) decode emotions from speech [*17-23*]. For example, Fairbanks and Pronovost [*24*] report correct identifications of simulated emotions to be in the 80 to 90% range, and Lieberman and Michaels [*18*] lend support for these values. As stated, however, much of these analyses were carried out on actors' voices, and thus, there is always the possibility that they artificially created those characteristics recognized by the auditors. However, the data reported by Williams and Stevens [*25*] suggest that such identifications can be validly made for emotions actually being experienced. In any case, if it is possible to identify accurately emotions from listening to the voice, it also should be possible to determine the relevant acoustical and temporal parameters within the voice signal that correlate with these states.

Speaking fundamental frequency (SFF or f0) is one such parameter. Fairbanks and Pronovost [*24*] suggest that the SFF level is increased for the emotions of fear and anger; Williams and Stevens [*8*] and Scherer [*26,27*] essentially agree. Further, Williams and Stevens [*25*] report that pilots and control tower operators showed an increase in SFF as a function of increased levels of stress, and this finding essentially is confirmed by Kuroda et al. [*10*]. On the other hand, not all authors agree with these observations. For example, Hicks [*16*] reports only slight increases in f0 as a function of stress; Markel et al. [*19*] did not observe perceived pitch to correlate with anger; and both Hecker et al. [*28*] and Almeida et al. [*29*] report that, although most of their subjects raised SFF in response to stress, others lowered it. Thus, although it is possible that increases in f0 level ordinarily accompany stress states, it is necessary to determine both the extent of a shift and its direction on an individual basis.

Vocal intensity is another acoustic parameter that can be measured and correlated with stress. However, only Hicks [*16*] has reported measurements of absolute intensity; he reports that stress correlated with increases in this parameter. Costanzo et al. [*30*] agree (at least with respect to loudness), as do Scherer [*20,27*] and Williams and Stevens [*25*]. Friedhoff et al. [*31*] also support this notion; they report that when they analyzed the speech of subjects who were lying (but said to be experiencing stress), vocal intensity increased as a function of task. On the other hand, Hecker et al. [*28*] did not find systematic (loudness) shifts to correlate with stress. In general, however, the evidence suggests that speech produced during stress probably is typified by increases in vocal intensity—at least for most people.

The prosodic analysis of a voice reflecting stress is a relatively complex process. For example, anger appears to be accompanied by rapid speaking rates as well as short phonatory durations and pauses [*32*]. Scherer [*20,26*] agrees, as do Markel et al. [*19*] with respect to rate, and Williams and Stevens [*25*] with respect to patterning. Further, Fairbanks and Hoaglin [*32*] contend that fear can be similarly characterized; Bachrach [*33*] and Scherer [*20,27*] agree, whereas Williams and Stevens [*25*] and Hicks [*16*] do not. A pattern reported by Hicks which appears to be generally accepted is that speaking during stress results in fewer speech bursts and that angry/fearful individuals appear to speak in longer utterances than they would ordinarily. Finally, Hicks [*16*] noted that nonfluencies appear to accompany stress. Except for Silverman and Silverman [*34*], all other authors tend to agree.

Finally, we organized a series of experiments and position papers [*16,35–39*]. Most were structured to test various elements related to the model summarized in Fig. 1. As can be seen, the voice of a stressed individual tends to change: speaking fundamental frequency rises; vocal intensity increases; speech rate increases slightly; nonfluencies can be observed; and the number of speech bursts is reduced. Our research tended to support this model, but

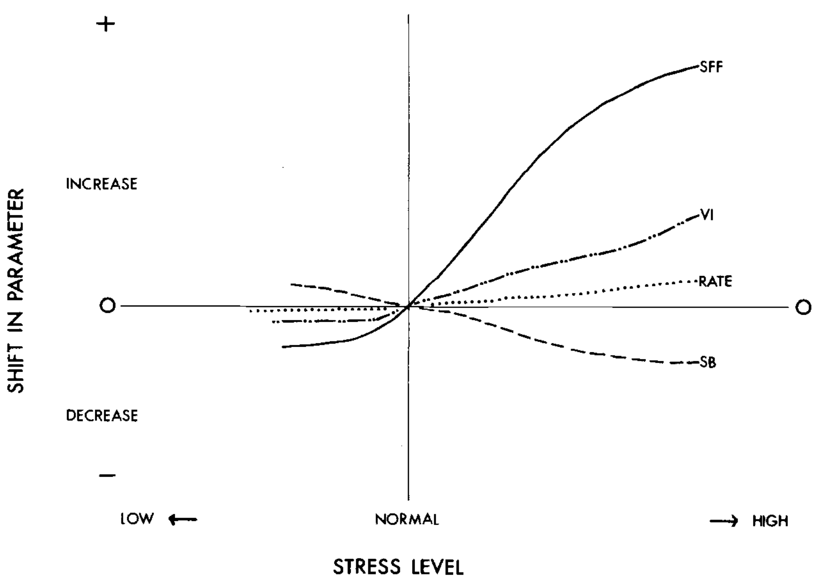

FIG. 1—*Shift in voice parameters as a function of stress level. Speaking fundamental frequency (SFF), vocal intensity (VI), and speaking rate all tend to increase with rise in stress, whereas the number of speech bursts (SB) tends to decrease.*

some deviations were noted. Moreover, since the observed characteristics occurred as shifts from normal or neutral speech states, they can be of value in the construction of a relevant *forensic* model only to the extent data on a given individual can be referenced to profiles based on that person's normal speech.

Second, the stress/lie relationships and the vocal correlates of deception issues must be addressed. While relatively little research has been carried out on these relationships, it often is assumed that lying and stress co-vary. However, Lykken [*4*] disagrees with this postulation. He maintains that stress can be induced when lying is not present and that the type of lie, especially with respect to the jeopardy involved, may differentially effect stress behavior. Further, and perhaps more seriously, there appear to be individuals (such as sociopaths) who may not exhibit a stress/lying correlation. The problem is further complicated by the fact that the exact nature and severity of these relationships are not known. Thus, although logic suggests that stress and lying may be related in some fundamental manner, the extent to which they are cannot be predicted.

Some research has been carried out directly on the vocal correlates of deception. For example, several authors indicate that SFF rises as a function of lying [*31,40-42*], but these shifts often are not statistically significant or they are "mixed" in nature. In addition, changes in other speech features, such as hesitations [*42,43*] and the patterns of speech bursts [*43*], are thought to accompany spoken deception. Thus, while generalizations are hazardous, especially since data are sparse, it must be conceded that there appear to be at least a few vocal correlates of lying, and that certain of the observed behaviors are similar to those known to accompany stress. Nevertheless, these stress/lie relationships are as yet not understood well enough to provide a theoretical model for the vocal analysis of deception.

The Voice Analyzers

Presumed Theoretical Bases

Supporters of the voice stress approach do not use the cited research to provide theoretical bases for their method. Moreover, their explanations are not articulated in any one place or in any great detail [*5,6*]. For example, relevant constructs can be found outlined (usually briefly) in the media, but the thrust there usually is one of sensationalism rather than clarification [*44-51*]. In any case, when psychological stress evaluator (PSE) operators are interviewed about their procedure, they ordinarily cite "microtremors" (see below) as the theoretical basis for their approach, but then shift to discussions of their "results." However, media reports such as these (or operating manuals) provide little insight relative to either the theoretical or scientific basis for the method.

A second source of information results from analyses of the equipment sold by the manufacturers of these devices. That is, it would appear that determination of their nature would provide information from which to infer the basis of the technique. However, our analysis of the PSE circuits (including those embedded in hard plastic) reveals that their typical operational mode probably parallels that of a simple low-pass filter. Indeed, VanDercar et al. [*52*] suggest that, except for the input and readout subsystems, the entire device consists only of a simple resistor-capacitor (RC) circuit, one that could be built for a few dollars. Hence, although we cannot deduce either the exact operation or the theory from circuit analysis alone, these inquiries permit some insight about the method being investigated.

When all sources are considered, the voice analysis proponents appear to base their technique on microtremors, that is, on the very slight tremblings that are thought to occur in some of the muscles of the human body. Although Horvath [*53*] provides an excellent discussion of this approach, a short review here would appear useful. To be specific, the voice analysis proponents contend that they are able to evaluate stress and lying by demodulating the "subsonic frequencies" that are caused by these microtremors and specifically by the minute oscillations found in the muscles of the vocal mechanism. They argue that these microtremors are normal to *any* voluntary muscle activity, but that when a person is stressed, they are suppressed. Thus, the PSE and similar devices presumably measure both the irregular modulation patterns in the signal trace of a normal vocal utterance and this same variability, which is reduced (or disappears), for stress [*6, 9, 54*]. Of course, there is no question that muscle microtremors exist, at least in the long muscles of the human extremities, and at rates varying from 8 to 14 kHz [*55, 56*]. However, the voice analyzer supporters apparently claim that such tremors also exist in the "voice"; they presumably are created by some interaction between the laryngeal muscles and the airstream. It is argued that these "vocal" tremors are the ones processed by their instrument and differentially identified in its graphic output. As would be expected, this explanation triggers many questions as to its validity.

First, Shipp and McGlone [*57*] tested the possibility that microtremors are present in the muscles associated with the vocal tract by embedding hook-wire electrodes in the lip and laryngeal muscles. They report that these muscles did not show tremor patterns similar to those of the long muscles, and Shipp and Izdebski [*58*] later confirm their position. They are forced to conclude then, that if the voice analyzers work at all, their operation has to be based on some other set of principles. McGlone [*59*] also agrees, arguing that tremors of the type in question usually are associated only with isometric contractions of the long muscles and seldom (if ever) occur in the small, fast-acting muscles of the vocal tract. Further, even if a microtremor could be detected, the muscle groups of the larynx and vocal tract operate in a manner so complex that operations as simple as a tremor would be mitigated. That is, as voice and speech are produced, any group of these muscles might contract and relax many times within a relatively short period of time—and in concert with, or opposition to, many other muscle groups. These rapidly shifting relationships clearly would mask any activity of the type discussed; they would do so, especially, as workload varied.

In any case, as the work of Eisenberg and Hill [*60*] also demonstrates, the actual process of muscle contraction is substantially more complex than are the concepts favored by the voice stress evaluators. Yet, Inbar et al. [*61*] claim to have been able to observe a laryngeal tremor of the type specified. Unfortunately, they employed surface electrodes coupled to a low-pass filter system. As is well known, surface electrodes will pick up activity created by any muscles in their vicinity. Thus, there is no way of knowing if the type of signals these authors report are from a single muscle, a group of muscles, interaction among muscles, or simply some sort of laryngeal "noise." Moreover, the research by Faaborg-Anderson [*62*] tends to contradict Inbar. He noted that firing rates of the vocalis muscle rose from 10 to 40 per second at the onset of phonation and remained there until after the cessation of sound.

Finally, attempts have been made to locate evidence of tremors by analysis of the acoustic signal. In 1976, McGlone and Hollien [*63*] report that they studied (spectrographically) the 5 to 100-Hz frequency band for subjects producing both stressed (random electric shocks) and unstressed speech. They found no evidence of any energy bands existing at frequencies below subjects' speaking fundamental frequency levels for either group. Almeida et al. [*29*] report similar data for the vocal tract, but Inbar et al. [*61*] claim to have observed acoustic evidence of a "tremor" occurring in the third formant of vowels. As can be seen, this latter statement both confounds the controversy and contradicts accepted acoustic theory relative to the interface between source function and vocal tract operation.

Perhaps a more telling argument involves the telephone (often used as part of this system of voice analysis): if a microtremor exists in the 8- to 14-Hz region, and the lower frequency of the telephone passband is 250 Hz (at best), some other feature (if any) must be the one that is being measured. Thus, it appears that none of the conditions established by Papcun [*64*]—that is, microtremors must occur in the vocal muscles, be manifest in the speech signal, and modified by stress—are supported by data. Bachrach [*33*] is forced to conclude "there appears to be no conclusive evidence that . . . a microtremor exists in the vocal apparatus . . . the transfer from normal physiological tremor to the vocal cords appears unwarranted."

Research Results

In a sense, it is immaterial how a system operates if it can perform the tasks required of it. Unfortunately, the available research does not appear to support the claims of the voice analysis proponents. Of course, there is some indication that these devices can detect stress if the level is extremely high [*63*], but it is ineffective if the stress level is somewhat lower [*59,65,66*]. Moreover, some contradictory data appear to exist. On the one hand, Brenner [*67*] reports that his PSE analyses showed talker stress to increase as a function of audience size, but these results are not always consistent with other of his data [*68*]. The findings of Leith et al. [*69*] add to the confusion. By PSE analysis, they found an apparent lowering of stress (that is, adaptation) as their subjects serially made telephone calls. However, the PSE did not discriminate between stutterers experiencing extremely high stress and controls (nonstutterers) in whom telephoning induced little-to-no stress. However, Brockway et al. [*70*] report that their obstetrical patients showed similar levels of stress on both the PSE and a standardized test of anxiety; Gorgen and Goodman [*71*] report PSE changes for a conflict task; and Smith [*72*] maintains that he can discriminate between phobics and other groups by application of this technique.

The VanDercar et al. [*52*] results are mixed, also. In their first study, these investigators report that PSE analyses correlated with stress measures of heart rate and scores on the State-Trait Anxiety Inventory (STAI). However, when the experiment was replicated under slightly different conditions, the PSE analyses did not reflect stress or correlate with the other two measures, and the Lynch and Henry [*73*] data are consistent with those for the second study. In short, it appears that this type of voice analysis sometimes provides an indication of the presence of stress in speech, but most often does not. Actually, however, research on stress is, in a sense, off the point because the primary objective of the voice analyzers is to detect deception from the person's utterances. Since such identification also is the goal of the members of the American Polygraph Society (APA), one would think that they would adopt or endorse these machines/techniques. Quite to the contrary, in 1973, the APA Board of Directors [*74*] voted against any use of these devices primarily because the "reliability and validity (of the technique) have not been demonstrated."

The experimental reports of detection of lying by voice stress analysis also are conflicting. For example, Heisse [*54*] reports having tested the PSE by requiring a group of 12 "trained" evaluators to process 258 "evaluation replies"; he claims they were correct 96.1% of the time. There may be some doubt relative to the rigor of Heisse's study because he appears to have had close ties with the manufacturer of PSE equipment, at least at that time.

On the other hand, both Barland [*65*] and Kubis [*8*], in independent experiments, have challenged the ability of the PSE to detect lying. Kubis' research appears pivotal; he used 174 subjects in an experiment involving a thief, a lookout, and an innocent suspect. He tested these subjects by means of a polygraph and two voice analyzers (PSE and voice stress analyzer [VSA]) operated by "trained" examiners. He also had the polygraph records reevaluated by independent examiners and obtained a subjective assessment by untrained tape monitoring personnel who were present during the interrogations. The accuracy levels were: polygraph, 76%; independent polygraph raters, 50 to 60; and monitors (subjective) scores, 50 to 60%. On the other hand, the PSE/VSA results were roughly at chance levels, and even when the "examiners" were provided with partial information, the obtained scores were well below those of the polygraphers. Kubis suggests that, because his untrained assistants were able to discriminate perceptually among his experimental subjects, he had demonstrated that the task resulted in sufficient induced emotionality to be valid for its stated purpose. Barland's [*56*] research is somewhat different in nature; he studied both low risk and high risk deception. He reports that PSE analysis was not sufficiently sensitive to detect lies if little jeopardy was involved; however, the results of his "high risk" experiment were mixed. Although Barland's results are not clear-cut enough to resolve the issue, they do little to support the claims of the voice analysis proponents.

Other investigators have reported stress analysis data also. Brenner and his associates [*68, 75*] report that when they offered their subjects a reward if they could "fool" the interrogators, most (subjects) were able to do so as the PSE analysis "failed to identify correct responses beyond chance levels" [*68*]. Moreover, they further criticize these devices on the basis of: scoring subjectivity, variability as a result of the utterance chosen for analysis, variation in recording quality, interpretation differences induced by variations in transcription speed, and the potential of subjects to control consciously their responses [*75*]. Data reported by Nachshon and Feldman [*76*] on the detection of concealed information tend to support the negative conclusions usually offered, as do those of Horvath [*77, 78*]. It is of little wonder, then, that Brenner and Branscomb [*75*] conclude that "there is now enough technical evidence to seriously question the PSE as a practical lie detection device." While this assessment appears accurate, it is obvious that more data are needed.

New Experiments

While the preponderance of the experimental data would argue that voice stress evaluation is not a valid technique, either for the detection of stress or deception, the fact remains that positive relationships sometimes are reported. Of course, it is possible that some of these data were biased toward a particular relationship or that the research design did not permit adequate testing. Nonetheless, some of the reported results are a little difficult to explain and a number of relationships need to be clarified. The experiments to follow are offered as contributions in this regard.

Detection of Stress Experiment

Purpose—The goal of the first experiment was to see if severe stress states could be detected by voice stress analysis. It was our judgment that, to be fair, the stress level should be both very high and verifiable by other techniques; that is, by procedures that are well documented as being both valid and reliable. Moreover, since it often is not possible to predict which stressor will affect a particular individual and by how much, we required that two different stress inducing situations be employed.

Subjects—Two populations were utilized in this research. The first consisted of seven males and five females; the second of ten males and seven females. All were healthy college students, between twenty and twenty-five years of age, who exhibited normal speech and hearing. All subjects were studied under two conditions: low (or no) stress and very high stress. The low stress condition was determined by subject ratings, experimenter evaluations, and a score of seven or less on the Multiple Affect Adjective Checklist (MAACL) anxiety scale [*79*]. To meet the high stress criteria, subjects had to be appropriately ranked on self and experimenter rating scales plus have received an anxiety score of thirteen or more on the MAACL.

Induced Stress—As stated, two different stressors were employed: electric shock with the first group (laboratory stress) and a first-time public speech for the second (situational stress). The first of the two subpopulations read a modified version of R. L. Stevenson's "An Apology for Idlers" twice; initially under the low stress condition; then while the subject received random shocks of up to 2.5 mA from a Grason-Stadler Psychogalvanometer. While this procedure did not harm the subject, it did cause enough discomfort to induce stress—and markedly so. The subjects for the second, or situational stress, procedure were drawn from a large number of volunteers who were to give their first presentation in a public speaking course. Only those tape recordings obtained from individuals who met the cited criteria were selected for analysis. In this case, the two speech samples were different in their nature, that is: public address—high stress; "Apology for Idlers"—low stress. However, they were of sufficient length (2 to 3 min) so that no contextual effects were judged present. Only 17 subjects (from the total of 51 volunteers) met the high-stress criteria and were included in this experiment.

Procedure—The procedure for all recordings was similar in nature. Subjects wore a specially constructed headset [*16*] with a small, calibrated microphone suspended on a boom. By this means, a constant microphone-to-lips distance of 8 in. (20 cm) could be maintained; a preexperimental familiarization period permitted subjects to become accustomed to this apparatus. Transmission was by FM radio-link to a tuner coupled to a SONY TC-353D tape recorder via a Hewlett-Packard 350D variable attenuator; all recordings were made in either a quiet classroom or a sound-treated room. They were checked for distortion or presence of noise; all proved acceptable. Finally, 20-s sections were drawn from the speech samples reflecting the lowest or highest stress, respectively. They were processed on a PSE-1 unit operated on the Mode III setting. Resulting segments were divided into four parts (four 5-s statements by each subject/condition), coded, mounted, and presented to the judges in a randomized sequence.

Judges—Three groups of auditors were employed. The first consisted of twelve college students who received a short training period (1 h) specifically based on the instructions found in the PSE manual [*6*]; they were not tested for competency. The second group consisted of ten college students who received yet further training; that is, over 2 h with a portion of the program consisting of supervised practice with samples drawn from the PSE manual. The third group of auditors consisted of five scientists (phoneticians with a minimum of two years' postdoctoral experience) who had carried out a substantial amount of research involving analysis of analog traces. This group also received the more extensive of the two training procedures.

Results—The results of this experiment have been presented in part orally [*80*]; the com-

TABLE 1—*Summary of evaluation of PSE traces of stressed and unstressed speech produced by two groups of talkers. Data are provided for three groups of auditors. All values are in percent and reflect the proportion of the time the auditors correctly identified the stress (or nonstress) state by this method.*

Auditor Groups Responses	Condition: Laboratory Stress, Stressed	Laboratory Stress, Unstressed	Situational Stress, Stressed	Situational Stress, Unstressed	Overall Performance
Student Group I (Minimum training):					
stressed	35	[a]	50	[a]	48
unstressed	[a]	45	[a]	64	
Student Group II (Full training):					
stressed	44	[a]	51	[a]	45
unstressed	[a]	45	[a]	39	
Scientist Group (Full training):					
stressed	45	[a]	48	[a]	52
unstressed	[a]	51	[a]	63	
Mean of Groups	41	47	50	55	(48)

[a]Reciprocal of other value.

plete data set may be found in Table 1. First, note that all scores hover around chance (that is, 50%). Since these data are based on over 6000 judgments, admittedly it is tempting to discuss what might appear to be "trends." For example, there may be a suggestion that professional training results in slightly better performance (assuming the technique is valid in the first place), but specific training with the PSE instrument does not. There also might be a tendency for subjects to be able to identify correctly unstressed speech at slightly higher levels than they can stressed speech. The scores for the situational stress certainly are better than those for the laboratory stress condition. However, it must be emphasized that the distributions were extremely variable and that no value even approached significance.

In sum, while it appears that application of the voice analysis technique occasionally results in the suggestion that stress is present if its level is extremely high, so can virtually *any* stress measurement technique if applied under like conditions. On the other hand, it does not appear that the voice stress method works very well in most situations. Worse yet, there is little to no indication as to what, if anything, is being measured. Perhaps the most important observation than can be made is that the use of the voice analysis method appears not to lead to correct behavioral classifications when stress is not present (or is very low) because many erroneous judgments of stress occurred under these conditions. Yet if non-stress speech is often identified as reflecting stress, a serious danger exists, that is, unfortunate interpretations will be made about the talker's integrity. It is clearly unacceptable that a substantial number of errors of this type occur.

The Detection-of-Deception Experiment

As is Horvath [*53*], we are sensitive to the voice analysis proponent's claim that their techniques are not adequately tested in the laboratory because only low risk lies can be induced in this milieu. For example, Heisse [*54*] argues that there are great differences between low-risk lies which he says cause the speaker little or no stress and high risk lies which cause a "great deal" of stress. He probably is correct to some extent. For example, certain authors [*65,66*] were not able to report a PSE output its manufacturers predict necessary if the procedure is to be viable in studies of low risk lies. In these instances, the lack of stress could have been attributed to the fact that the subject was lying not out of fear of some penalty but rather at the request of the experimenter. Nevertheless, we challenge the contention of the voice analysis proponents because, in our opinion, their argument is not a valid one. Since the intensity of stress associated with deception obviously varies from situation to situation and from subject to subject, a system that validly detects deception should be able to do so under virtually any or all circumstances. If it does not, it simply cannot be considered valid. However, to counter this potential criticism, we carried out the experiment described below.

As stated, voice stress evaluators are said not to assess lying directly, but, rather, to respond to stress in speech. Accordingly, we judged that a fair evaluation could be carried out if we required the experimental subjects to produce utterances which they knew were untrue and which were particularly stressful to them. The factor chosen as motivation for these "high risk" lies (that is, the stressor) was the talker's sense of honor. It was judged that, if subjects were convinced that an audience important to them was to hear them espousing a view which was strongly in conflict with their true feelings, the resulting speech sample would reflect lying with substantial jeopardy. Accordingly, a group of volunteers were identified, all of whom were *strong* antivivisectionists. Before the experiment, one of the investigators explained the seriousness and importance of the project and in particular led the potential subjects to believe that auditors from the "University and Gainesville communities" would hear their statements. They further were led to believe that the purpose of the research was to see if the cited auditors could discriminate the "lie" statements from the truth. Subjects were not told until after the entire experiment had been completed that the auditors would not actually *hear* the speech samples but rather "read" (coded) traces produced by a

PSE device, and that neither their identity or the speech sample was compromised in any manner.

Subjects—In all, 44 individuals volunteered to be subjects. As part of the selection procedure, they first were required to view a series of visual materials which showed animals being severely abused in experiments. They were then asked to read and tape record two 20-s statements, one strongly antivisection and one strongly in support of animal experiments under any and all conditions. Three evaluations of the stress levels of the subjects were employed: self reports; experimenter observations; and scores on the MAACL test [*79*]. The experimental and control groups were chosen on the basis of all three evaluations; if an individual did not meet any of them, he or she was eliminated from the subject pool. To be a member of the experimental group, a subject had to score between 1 and 8 after the low stress (truthful) reading and between 10 and 21 in conjunction with the high stress (lie) passage—with a minimum difference of four points between the two scores—plus meet the subjective criteria; twelve individuals did so. Conversely, to be a member of the control group, a volunteer had to score between 7 and 12 on the MAACL for both conditions, with no more than a single point between tests, and these conditions had to be confirmed by the subjective tests; seven individuals qualified for the control group. Most of the subjects who did not qualify for either group (N = 19) did not do so because both of their scores were very high, presumably because the overall task simply was too difficult for them. In any case, the experimental groups showed a sharp contrast between lying with jeopardy (high stress patterns) and speaking a truth with low stress.

Procedure—All speech samples were recorded in a sound-treated room (IAC 1600) and on laboratory quality equipment. The recordings were processed on a PSE-1 operated on Mode III. The resulting traces were mounted (one statement per card; eight samples per subject), coded, and randomized.

Judges—Three groups of examiners were used as judges; they included: ten college students who were "taught" to read PSE traces by means of the 2-h training program (they were different subjects than those used in the first experiment but trained similarly to the second group); five experienced scientists (the five used in the prior experiment); and three trained/experienced PSE operators who volunteered to assist with the research.

Results—The results of this experiment can be best understood by consideration of Table 2. As may be seen, the experimental group was judged to be lying less than 50% of the time

TABLE 2—*Summary of percent of correct deception judgments by three groups of auditors. Talkers were individuals who recorded a true statement under low stress and a deceptive one under high stress: controls were individuals who showed only moderate stress for both conditions.*

	Conditions			
	Experimental Group (N = 12)		Control Group (N = 7)	
Auditors	Deception	True	Deception	True
Students				
true	[a]	40	[a]	39
untrue	48	[a]	40	[a]
Phoneticians				
true	[a]	37	[a]	43
untrue	62	[a]	54	[a]
PSE Operators				
true	[a]	58	[a]	48
untrue	44	[a]	43	[a]
Mean	51	45	46	43

[a]Reciprocal of paired value.

by the students and PSE operators and only a little over 60% by the phoneticians when, in fact, they were lying 100% of the time; a chi square statistical procedure was carried out and this value (that is, that for the phoneticians) was the only one (of the twelve) that was significantly different from the rest at the 0.05 level of confidence. Actually, however, the obtained results essentially are at chance levels, and even the 58% correct truthful identifications by the PSE operators was not of significance. Further, it should be remembered also that the PSE operators concluded the subjects are lying—when, indeed, they are telling the truth—an incredible 42% of the time. Imagine the appalling disservice to clients and subjects that would occur in an instance such as this one! Nor do the data obtained on the control group offer much encouragement. In short, it is difficult to defend the validity of a system when various groups of examiners cannot discriminate among groups who are telling the truth under conditions of low stress, speaking a falsehood under high stress, or telling (sometimes) either the truth or a lie under moderate to low stress. If this method were valid, it would appear that more than one of the three groups of auditors would have scored significantly above chance on one or more of the conditions. Accordingly, it must be concluded that voice analyzers are not very good tools (if they are effective at all) for the detection of deception. Even if they did work, they are not in a class with the polygraph[3] [*8,53,59,65*]—and the limitations of the polygraph are well known. Finally, no one as yet has tested these devices after the speech sample has been transmitted over a telephone—research of this type is very much needed.

To conclude, although the proponents of voice analyzers may argue that they are effective as stress and lie detectors, data are available that support a far stronger case to the contrary. Negative arguments about this approach become much more compelling when it is remembered that individuals who make claims about a device, any device, must demonstrate the validity of their contentions. Perhaps even more serious are the well-documented fears that the use and misuse of these units will lead to abuses. Finally, even though scientists working in this area are keenly aware of the substantial need (by law enforcement agencies) for a valid tool of this type, the available research argues against their use. In short, the inescapable conclusion is that a device that will permit the detection of lying from the voice analysis alone presently does not exist.

Acknowledgments

This project is dedicated to the late Robert E. McGlone.

References

[*1*] Reid, J. E. and Inbau, F. E., *Truth and Deception: The Polygraphy ("Lie Detector") Technique*, Williams and Wilkins, Baltimore, 1977.

[*2*] Matte, J. A., *The Art and Science of the Polygraph Technique*, Charles C Thomas, Springfield, IL, 1980.

[*3*] Lykken, D., "Psychology and the Lie Detector Industry," *American Psychologist*, 1974, pp. 725-739.

[*4*] Lykken, D., *A Tremor in the Blood*, McGraw-Hill, New York, 1981.

[*5*] Bennett, R. H., Jr., *Hagoth: Fundamentals of Voice Stress Analysis*, Hagoth Corp., Issaquah, WA, 1977.

[*6*] Anonymous, *PSE Manual*, Dektor, Counterintelligence and Security, Inc., Springfield, VA (not dated).

[*7*] Geison, L. L., *Evaluation of High Stress Lying by Voice Analysis*, unpublished M.A. thesis, University of Florida, 1979.

[*8*] Kubis, J., *Comparison of Voice Analysis and Polygraph as Lie Detection Procedures*, Technical Report LWL-CR-U3B70, U.S. Army Land Warfare Laboratory, Aberdeen Proving Ground, MD, 1973.

[3] T. Puckett, personal communication relative to unpublished study, "Voice Stress Analysis Procedures Vis-a-Vis the Polygraph Procedure in Real Life Testing Situations," 1981.

[9] Kupec, E. W., "Truth or the Consequences," *Law Enforcement Community*, Vol. 4, 1977, pp. 12-18, 42-45.
[10] Kuroda, I., Fujiwara, O., Okamura, N., and Utsuki, N., "Method for Determining Pilot Stress Through Analysis of Voice Communication," *Aviation, Space and Environmental Medicine*, Vol. 47, 1976, pp. 528-533.
[11] Williams, C. E. and Stevens, K. N., "On Determining the Emotional State of Pilots During Flight: An Exploratory Study," *Aerospace Medicine*, Vol. 40, 1969, pp. 1369-1372.
[12] Appley, M. H. and Trumbull, R., "On the Concept of Psychological Stress" in *Psychological Stress: Issues in Research*, M. H. Appley and R. Trumbull, Eds., Meredith, New York, 1967.
[13] Lazarus, R. S., *Psychological Stress and the Coping Process*, McGraw-Hill, New York, 1966.
[14] Scherer, K. R., "Vocal Indicators of Stress" in *Speech Evaluation in Psychiatry*, J. Darby, Ed., Grune and Stratton, New York, 1981, pp. 171-187.
[15] Basowitz, H., Persky, W., Korchin, J. T., and Grinter, R. R., *Anxiety and Stress*, McGraw-Hill, New York, 1955.
[16] Hicks, J. W., Jr., "An *Acoustical/Temporal Analysis of* Emotional Stress in *Speech*," unpublished Ph.D. dissertation, University of Florida, 1979.
[17] Davitz, J. R. and Davitz, L. J., "The Communication of Feelings Through Content-Free Speech," *Journal of Communication*, Vol. 9, 1959, pp. 6-13.
[18] Lieberman, P. and Michaels, S. B., "Some Aspects of Fundamental Frequency and Envelope Amplitude as Related to Emotional Content of Speech," *Journal of the Acoustical Society of America*, Vol. 34, 1962, pp. 922-927.
[19] Markel, N. N., Bein, M. F., and Phillis, J. A., "The Relationship Between Words and Tone of Voice," *Language and Speech*, Vol. 16, 1973, pp. 15-21.
[20] Scherer, K. R., "Acoustic Concomitants of Emotional Dimensions: Judging Affect from Synthesized Tone Sequences" in *Nonverbal Communication*, S. Weitz Ed., Oxford University Press, New York, 1974, pp. 249-253.
[21] Scherer, K. R. and Oshinsky, J. S., "Cue Utilization in Emotion Attribution from Auditory Stimuli," *Motivation and Emotion*, Vol. 1, 1977, pp. 331-346.
[22] Simonov, P. V., Frolov, M. V., and Tabgkin, V. L., "Use of the Invariant Method of Speech Analysis to Discern the Emotional State of Announcers," *Aviation, Space, and Environmental Medicine*, Vol. 46, 1975, pp. 1014-1016.
[23] Starkweather, J. A., "Vocal Communications of Personality and Human Feelings," *Journal of Communications*, Vol. 11, 1961, pp. 63-72.
[24] Fairbanks, G. and Pronovost, W., "An Experimental Study of the Pitch Characteristics of the Voice During the Expression of Emotion," *Speech Monographs*, Vol. 6, 1939, pp. 87-104.
[25] Williams, C. E. and Stevens, K. N., "Emotions and Speech: Some Acoustical Correlates," *Journal of the Acoustical Society of America*, Vol. 52, 1972, pp. 1238-1250.
[26] Scherer, K. R., "Effects of Stress on Fundamental Frequency of the Voice," *Journal of the Acoustical Society of America*, Vol. 62, 1977, pp. S25-26(A).
[27] Scherer, K. R., "Personality Markers in Speech" in *Social Markers in Speech*, K. R. Scherer and H. Giles, Eds., Cambridge University Press, Cambridge, U.K., 1979, pp. 147-209.
[28] Hecker, M. H. L., Stevens, K. N., von Bismarck, G., and Williams, C. E., "Manifestations of Task-Induced Stress in the Acoustic Speech Signal," *Journal of the Acoustical Society of America*, Vol. 44, 1968, pp. 993-1001.
[29] Almeida, A., Fleischmann, G., Heike, G., and Thormann, E., "Short Time Statistics of the Fundamental Tone in Verbal Utterances Under Psychic Stress," presented at the 8th International Congress of Phonetic Sciences, Leeds, England, 1975.
[30] Costanzo, F. S., Markel, N. N., and Costanzo, P. R., "Voice Quality Profile and Perceived Emotion," *Journal of Counsel Psychology*, Vol. 16, 1969, pp. 267-270.
[31] Friedhoff, A. J., Alpert, M., and Kurtzberg, R. L., "An Electro-Acoustical Analysis of the Effects of Stress on Voice," *Journal of Neuropsychology*, Vol. 5, 1964, pp. 265-272.
[32] Fairbanks, G. and Hoaglin, L. W., "An Experimental Study of the Durational Characteristics of the Voice During the Expression of Emotion," *Speech Monographs*, Vol. 8, 1941, pp. 85-90.
[33] Bachrach, A. J., "Speech and Its Potential for Stress Monitoring" in *Proceedings, Workshop on Monitoring Vital Signs in the Diver*, C. E. G. Lundgren, Ed., Undersea Medical Society and Office of Naval Research, Washington, DC, 1979, pp. 78-93.
[34] Silverman, F. H. and Silverman, E. M., "Effects of Threat of Shock for Being Disfluent on Fluency of Normal Speakers," *Perceptual and Motor Skills*, Vol. 41, 1975, pp. 353-354.
[35] Hicks, J. W., Jr. and Hollien, H., "The Reflection of Stress in Voice-1: Understanding the Basic Correlates" in *Proceeding of the Carnahan Conference on Crime Countermeasures*, Lexington, KY, 1981, pp. 189-194.
[36] Hollien, H., "Vocal Indicators of Psychological Stress" Chapter in *Forensic Psychology and Psychiatry*, F. Wright, C. Bahn, and R. W. Rieber, Eds., New York Academy of Sciences, New York, 1980, pp. 47-72.

[37] Hollien, H., "Acoustic Correlates of Psychological Stress" in *Transcripts of the Tenth Symposium on Care of the Professional Voice*, V. Lawrence, Ed., The Voice Foundation, New York, 1981, pp. 145-158.
[38] Hollien, H., "Forensic Communication: An Emerging Specialty," *Criminal Defense*, Vol. 10, 1983, pp. 22-29.
[39] Hollien, H. and Darby, J. K., "Acoustic Comparisons of Psychotic and Non-Psychotic Voices" in *Current Issues in the Phonetic Sciences*, H. Hollien and P. A. Hollien, Eds., John Benjamins, B.V., Amsterdam, 1979, pp. 829-835.
[40] Ekman, P., Friesen, W. V., and Scherer, K. R., "Body Movement and Voice Pitch and Deceptive Interaction," *Semiotica*, Vol. 16, No. 1, 1976, pp. 23-27.
[41] Krauss, R. M., Geller, V., Olson, C., and Appel, W., "Pitch Changes During Attempted Deceptions," *Journal of Personality and Social Psychology*, Vol. 35, No. 5, 1977, pp. 345-350.
[42] Sawyer, C., "Detecting the Unspoken Deception," *Security World*, Jan. 1983, pp. 50-51.
[43] Hocking, J. E. and Leathers, D. G., "Nonverbal Indicators of Deception: A New Theoretical Perspective," *Communication Monographs*, Vol. 47, 1980, pp. 119-131.
[44] "Dallas: New Questions and Answers," *Newsweek*, 28 April 1975, pp. 36-38.
[45] Dick, W., "Scientific Evidence Proves Ted Told the Truth About Chappaquidick," *National Enquirer*, 1 July 1975, p. 49.
[46] O'Toole, G., "Lee Harvey Oswald Was Innocent," *Penthouse*, 1975, pp. 45-46; 124-132.
[47] Dworken, A., "Patty Hearst Not Guilty, Voice Test Proves She Was Forced to Lie," *National Enquirer*, 23 Sept. 1975, p. 50.
[48] Dick, W., "Sirhan Was Hypnotized to Kill Bobby Kennedy," *National Enquirer*, 27 Oct. 1975, p. 50.
[49] Haines, R., "Elizabeth Ray Told the Truth About the Washington Sex Scandal: Congressman Hayes Did Not," *National Enquirer*, April 1976, p. 51.
[50] Hart, C., "Scientific Truth Detector Reveals . . . Laren Bacall is Lying in her Ford TV Commercial," *National Enquirer*, 29 July 1980, p. 54.
[51] Cooke, J., "Scientific Truth Detector Reveals: Sinatra Told the Truth When He Denied Mob Ties," *National Enquirer*, Vol. 7, 20 March 1981.
[52] 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.
[53] Horvath, F., "Detecting Deception: The Promise and Reality of Voice Stress Analysis," *Journal of Forensic Sciences*, Vol. 27, No. 2, April 1982, pp. 340-351.
[54] Heisse, J. W., "Audio Stress Analysis—A Validation and Reliability Study of the Psychological Stress Evaluator (PSE)" in *Proceedings of the Carnahan Conference on Crime Countermeasures*, Lexington, KY, 1976, pp. 5-18.
[55] Brumlik, J. and Yap, C., *Normal Tremor: A Comparative Study*, Charles C Thomas, Springfield, IL, 1970.
[56] Lippold, O., "Physiological Tremor," *Scientific American*, Vol. 224, 1971, pp. 65-73.
[57] Shipp, T. and McGlone, R. E., "Psysiologic Correlates of Acoustic Correlates of Psychological Stress," *Journal of the Acoustical Society of America*, Vol. 53, 1973, p. S63(A).
[58] Shipp, T. and Izdebski, K., "Current Evidence for the Existence of Laryngeal Macrotremor and Microtremor," *Journal of Forensic Sciences*, Vol. 26, No. 3, July 1981, pp. 501-505.
[59] McGlone, R. E., "Tests of the Psychological Stress Evaluator (PSE) as a Lie and Stress Detector" in *Proceedings of the Carnahan Conference on Crime Countermeasures*, Lexington, KY, 1975, pp. 83-86.
[60] Eisenberg, E. and Hill, T. L., "Muscle Contraction and Free Energy Transduction in Biological Systems," *Science*, Vol. 227, 1985, pp. 999-1006.
[61] Inbar, G. F., Eden, G., and Kaplan, M. A., "Frequency Modulation in the Human Voice and the Source of Its Mediation" in *Proceedings of the Carnahan Conference on Crime Countermeasures*, Lexington, KY, 1977, pp. 213-219.
[62] Faaborg-Anderson, K., "Electromyographic Investigation of Intrinsic Laryngeal Muscles in Humans," *Acta Physiologica Scandinavia*, Vol. 41 (Supplement), 1957, p. 140.
[63] McGlone, R. E. and Hollien, H., "Partial Analysis of the Acoustic Signal of Stressed and Unstressed Speech" in *Proceedings of the Carnahan Conference on Crime Countermeasures*, Lexington, KY, 1976, pp. 19-21.
[64] Papcun, G., "The Effects of Psychological Stress on Speech," paper delivered to the Fall meeting of the Acoustical Society of America, Los Angeles, CA, 1973.
[65] Barland, G., "Use of Voice Changes in Detection of Deception," *Journal of the Acoustical Society of America*, Vol. 54, 1973, p. 63(A).
[66] McGlone, R. E., Petrie, C., and Frye, J., "Acoustic Analysis of Low-Risk Lies," *Journal of the Acoustical Society of America*, Vol. 55, 1974, p. S20(A).

[67] Brenner, M., "Stagefright and Steven's Law," paper presented at the Eastern Psychological Association Meetings, April 1974.
[68] Brenner, M., Branscomb, H. H., and Schwartz, G. E., "Psychological Stress Evaluator—Two Tests of a Vocal Measure," *Psychophysiology*, Vol. 16, 1979, pp. 351-357.
[69] Leith, W. R., Timmons, J. L., and Sugarman, M. D., "The Use of the Psychological Stress Evaluator with Stutterers," paper read to the Convention of the American Speech and Hearing Association, 1977.
[70] Brockway, B. F., Plummer, O. B., and Lowe, B. M., "The Effects of Two Types of Nursing Reassurance Upon Patient Vocal Stress Levels as Measured by a New Tool, the PSE," *Nursing Research*, Vol. 25, 1976, pp. 440-446.
[71] Borgen, L. and Goodman, L., "Voiceprint Analysis of Anxiolytic Drug Effects: Preliminary Results," *Clinical Pharmacology and Therapeutics*, Vol. 19, 1976, p. 104.
[72] Smith, G. A., "Voice Analysis for the Measurement of Anxiety," *British Journal of Medical Psychology*, Vol. 50, 1977, pp. 367-373.
[73] Lynch, B. and Henry, D., "A Validity Study of the Psychological Stress Evaluator," *Canadian Journal of Behavioral Science*, Vol. 11, 1979, pp. 89-95.
[74] Board of Directors, "Resolution Concerning the Dektor Psychological Stress Evaluator (PSE-1)," American Polygraph Association, Miami, FL, 1973.
[75] Brenner, M. and Branscomb, H. H., "The Psychological Stress Evaluator, Technical Limitations Affecting Lie Detection," *Polygraph*, Vol. 8, 1979, pp. 127-132.
[76] Nachshon, I. and Feldman, B., "Vocal Indices of The Psychological Stress Evaluator," *Journal of Police Science and Administration*, Vol. 8, 1980, pp. 40-53.
[77] 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, 1978, pp. 338-344.
[78] 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, 1979, pp. 323-330.
[79] Zuckerman, M., Lubin, B., Vogel, L., and Valeruis, E., "Measurement of Experimentally Induced Affects," *Journal of Consulting Psychology*, Vol. 28, 1964, pp. 418-425.
[80] Hollien, H., Geison, L. L., and Hicks, J. W., Jr., "Stress/Lie Studies Utilizing the PSE," paper read at the Annual Meeting of the American Academy of Forensic Sciences, New Orleans, LA, 1980.

Address requests for reprints or additional information to
Harry Hollien, Ph.D.
ASB-50, IASCP
University of Florida
Gainesville, FL 32611