ANALYSIS OF THE VOICE

A Study

by

G. Alan Smith
Department of Psychology
Powick Hospital
Worcester, Endland

1973

The study investigated a new technique by which stress is detected by the electronic analysis of tape recordings of a speaker. Patients read out phrases referring to common life stressors and also stressors peculiar to the individual patient. The anxiety caused to the patient by these stressors was assessed from the electronic analysis of his voice, and there was evidence that this could be done with an adequate degree of reliability. There was evidence also of some degree of validity, as the voice evaluations in general corresponded significantly with the patients' self-reports about their anxieties. Moreover, this correspondence was low only for people with high Lie scores on the Eysenck Personality Inventory, people whose self-reports might be expected to have least validity. It is suggested that psychological methods of treatment might be made more efficient by using this technique to more accurately identify areas of anxiety.

PSYCHOLOGICAL STRESS EVALUATOR

This invention relates to the detection and measurement of psychological stress and, more particularly, to the measurement of psychological stress at levels below those discernible without instrumentation and the measurement of psychological stress without the use of applied or visible sensors.

The existence of the relationships between psychological states and physiological responses to such states has long been recognized. Two general types of exploitation of these relationships have developed. The first of these is the causation of programmed response to psychological stimuli, which has been most notably documented by Pavlov in his experiments with conditioned responses, and has provided some basis for such divergent fields as applied psychology, advertising, and hypnotism. This type of exploitation of the psycho-physiological relationships is only supplementally involved with the present invention.

The second general type of exploitation of the psycho-physiological relationships involves the identification and measurement of psychological changes by identifying and measuring resultant physiological changes. This approach is used largely in medical and psychological research and diagnosis, and in lie detection. Two methods of evaluation are used. When the responses are sufficiently great, some subjective evaluations can be made by a trained observer; however, far more accurate evaluations can be made by instrumentation designed to detect and measure relatively small degrees of physiological change. Those physiological changes most usually considered are brainwave patterns, heart activity, skin conductivity, and the breathing activity. While the measurement of these activities does provide a far more accurate evaluation of physiological response than does direct observation, it introduces several disadvantageous features. The most functionally serious of these problems is the artificiality of the testing situation caused by the requirement for sensors to be attached to the subject of the examination, the controlled environment usually required, and resultant restrictions on the normal activity of the subject. These requirements can be expected to induce a psychological set in the subject which, in some cases, may be as strong as or stronger than the psychological set which is to be evaluated, thereby substantially reducing the validity of the evaluation

It has been an objective of this invention to achieve a method for psychological stress evaluation which avoids the testing and evaluation constraints imposed by the existing requirements for obvious and attached sensors.

It has been a specific objective of this invention to use the human voice as the medium which provides the physiological response to psychological change -- inasmuch as the use of this medium resolves the existing problems of attached sensors, elaborate equipment, and highly controlled environment in association with the subject; inasmuch as this is the one physiological medium which can be transmitted over distances by existing and readily available techniques (telephone and radio); and inasmuch as this is the one physiological medium which can be recorded simply as a completely valid and comprehensive analog of the medium (e.g., by tape recording).

The basic physiology of the human voice, as heretofore understood, produces two types of sound. The first of these is a product of the vibration of the vocal chords, which, in turn, is a product of partially closing the glottis and forcing air through the glottis by contraction of the lung cavity and the lungs. The frequency of these vibrations may vary generally between 100 and 300 Hertz, depending upon the sex and age of the speaker and upon the intonations the speaker applies. This sound has a rapid decay time.

The second type of sound involves the formant frequencies. This is sound which results from the resonance of the cavities in the head (throat, mouth, nose, and sinus cavities). This sound is created by excitation of the resonant cavities by a sound source of lower frequencies, in the case of the vocalized sound produced by the vocal chords, or b y the partial restriction of the passage of air from the lungs, as in the case of unvoiced fricatives. Whichever the excitation source, the frequency of the formant is determined by the resonant frequency of the cavity involved. The formant frequencies appear generally above 800 Hertz and appear in distinct frequency bands which correspond to the resonant frequency of the individual cavities. The first, or lowest, formant is that created by the mouth and throat cavities and is notable for its frequency shift as the mouth changes its dimensions and volume in the formation of various sounds, particularly vowel sounds. The higher formant frequencies are more constant, because of the more constant physical volume of the cavities. The formant waveforms are ringing signals, as opposed to the rapid decay signals of the vocal chords. When voiced sounds are uttered, the voiced waveforms are imposed upon the formant waveforms as amplitude modulation.

A third signal category in the human voice has been uncovered during the research leading to this invention. This is an infrasonic frequency modulation, present, in some degree, in both the vocal chord sounds and in the formant sounds. This signal is typically between 8 and 12 Hertz; therefore, it is inaudible to the human ear. As frequency modulation, as opposed to amplitude modulation, it is not directly discernible on time base/amplitude chart recordings. Because this infrasonic signal is one of the more significant voice indicators of psychological stress, it will be dealt with in greater detail.

There are several analogies which are used to provide schematic representations of the entire voice process. Both mechanical and electronic analogies are successfully employed, for example, in the design of computer voices. These analogies, however, consider the voiced sound source (vocal chords) and the walls of the cavities as hard and constant features. As opposed to these analogies, both the vocal chords and the walls of the major formant-producing cavities are soft flexible tissue immediately responsive to the complex array of muscles which control them. Those muscles which control the vocal chords thru the mechanical linkage of bone and cartilage allow both the purposeful and automatic production of voiced sound and variation of voice pitch. Similarly, those muscles which control the tongue, lips, and throat allow both the purposful and the automatic control of the first formant frequencies. Other formants can be affected similarly to a more limited degree. It is worthy of note that, during normal speech, these muscles are performing at a small percentage of their work capability. For this reason, in spite of their being employed to change the position of the vocal chords and the position of the lips, tongue, and inner throat walls, the muscles remain in a relatively relaxed state. As a result of research conducted for this invention, it has been determined that during this relatively relaxed state a natural muscular undulation occurs at the typically 8 to 12 Hertz frequency previously mentioned. This undulation slightly varies the tension of the vocal chords and causes audibly indiscernible shifts in the basic pitch frequency of the voice and it slightly varies the volume of the resonant cavity (particularly that associated with the first formant) and the elasticity of the cavity walls to cause audibly indiscernible shifts in the formant frequencies. These shifts about a central frequency constitute a frequency modulation of the central or carrier frequency. In order to observe this frequency modulation, any of the several existing techniques for the demodulation of frequency modulation may be employed, bearing in mind, of course, that the modulation frequency is the nominal 8 to 12 Hertz and the carrier is one of the bands within the voice spectrum. Figure 1 depicts pulses of the amplitude modulation of formants by a voiced signal of 190 Hertz, resulting from 2.5 Kilohertz high-pass filtering of the voice of an unstressed subject and the contained infrasonic frequency modulation observable in the center-of-mass waveform. A simple slope detection was made of the unlimited infrasonic waveform of a syllable utterance.

It has been determined, further, that the array of muscles associated with the vocal chords and the cavity walls are subject to mild muscular tension when slight to moderate psychological stress is created in the subject. This tension, indiscernible to the subject, is sufficient to decrease or virtually eliminate the muscular undulations present in the unstressed subject, thereby removing the basis for the carrier frequency variations which produce the infrasonic frequency modulation. Figure 2 depicts a similar utterance as that of Figure 1, but at a time of induced psychological stress. In this case the center-of-mass waveform can be seen to be essentially devoid of the infrasonic variations observed in the unstressed utterance in Figure 1, even though all other test factors and the demodulation procedures were held constant.

While the use of the infrasonic waveform is unique to the use of voice as the physiological medium for psychological stress evaluation, the voice provides for additional instrumented indications of aurally indiscernible physiological changes as a result of psychological stress, which physiological changes are detectable also by current means. Of the four most usually considered physiological changes previously mentioned (brainwave patterns, heart activity, skin conductivity, and breathing activity), two – breathing activity and heart activity – directly and indirectly affect the amplitude and the detail of an utterance waveform and provide the basis for a more gross evaluation of psychological stress, particularly when the testing situation involves sequential vocal responses. Figure 3 depicts a test in which psychological stress was induced in a female subject by having her utter a hypothetical lie in a Peak of Tension test, in which she responded, "I did not choose that letter" for each of a series of 12 letters, one of which, in fact, she had selected. The chart displays an aggregate of pneumographic, cardiographic, and, to a lesser degree, infrasonic influence. The signal for this display has been highly integrated to show more clearly the gross aggregate effect. This technique is particularly useful in determining stress zones in longer statement runs, by avoiding waveform complexity. This chart, the 12 letter series from K to V, shows the following significant features which are indicative of physiological changes caused by the attempted deception.

There is a marked constriction at K which reflects the beginning-of-test tension normal in most overt testing situations. Normally, this initial tension would dissipate at L unless the set of the individual were such that the initial tension is reinforced by the tension of anticipating the approaching lie. Such is the case in this test, and tension is seen to continue at M, N, and O, the point of deception. In addition to the observable effects of the infrasonic signal (which is normally present in relaxed speech but attenuated or absent in stressed statements) there are certain more usual physiological indicators also present which portray the changing emotional pattern of the individual; these are supression/hyperventilation, diction stress, and significant delays in response time. The pneumographic influence gives the greatest indication of relief after the point of deception in that a noticeable increase of amplitude is indicated at the point where the individual passed the point of deception and began to compensate for the decrease in oxygen/carbon dioxide exchange occuring at K through O. Beginning at P the relief experienced by subject is evident through the remainder of the chart with the partial exception of Q, which can be expected as a momentary fear on the part of the subject that relief at P may have been evident audibly to the examiner.

Additionally, there are certain diction stresses which may be evident as a progressive change in an individual pattern. This may or may not be audible as subject exercises abnormal control over his diction in an attempt to maintain a static speech pattern. These indications include minor changes in individual syllable stress and changes in the concatenation patterns in the separate responses. These indicators are largely responsible for the waveform pattern (as distinguished from amplitude) of the individual response displays; they follow a slightly different progression to O, the point of deception, in that they are not as much involved with the beginning-of-test stress demonstrated by the lower amplitude at K, but cause an increasingly less complex display up to N and O, and suddenly return to their complexity with the marked psychological relief at P. With the exception of Q (for the reason previously discussed) this non-stressed pattern continues throughout the remainder of the test.

It should be noted that the opposite configuration may occur; that is, the stress may be indicated by high amplitude and a multi-form trace while relief may be shown by a drop in amplitude and a more simple pattern. This, of course, depends upon whether a given individual responds to a given psychological stimulus with excitation or depression. While the general indicators of stress and relief may differ from test to test, they are relatively stable within any individual test and, of course, the infrasonic indicator remains constant from test to test and from individual to individual.

Figure 4 depicts an expansion of the N-O-P portion of this Peak of Tension test to show the somewhat constrained infrasonic waveform at N, as the tension is increasing, the more greatly constrained infrasonic waveform at O (the "lie") and the presence of the infrasonic waveform at P, as a result of psychological relief.

Figure 5 depicts similar portions of two narrative-type utterances made by a subject under unstressed conditions (5a) and under conditions of induced mild anxiety (5b). As has been indicated in the previous examples, the infrasonic waveform is obvious in the unstressed utterance and is greatly attenuated in the stressed utterance.

While the above descriptions and examples deal with the psycho-physiological relationships from the standpoint of an overview, those who are knowledgeable in these areas will readily recognize the functions of the endocrine glands, and the sympathetic and parasympathetic nervous systems in completing the course between the psychological stimulus and the several physiological responses involved in this invention. Similarly, while the details of the physiology of the larynx and the resonant cavities of the throat and head have been described only to the point deemed necessary to support the techniques and findings, those persons versed in human physiology will be aware of the well-known physical features involved in these areas.

FIG. I

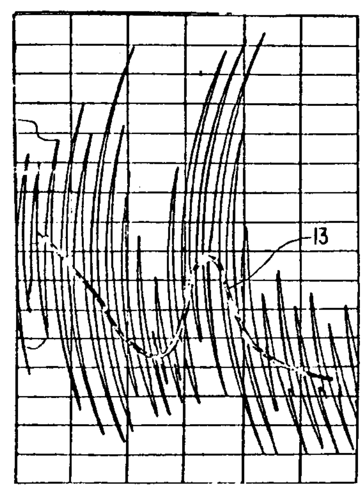
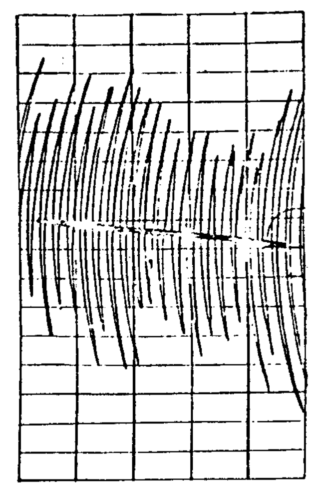


FIG. 2



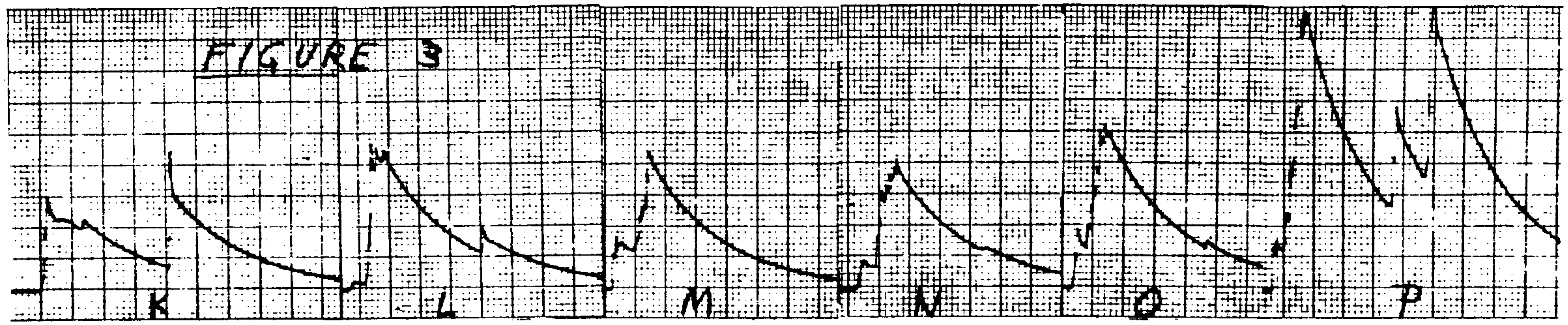


FIGURE 3
K
L
M
N
O
P

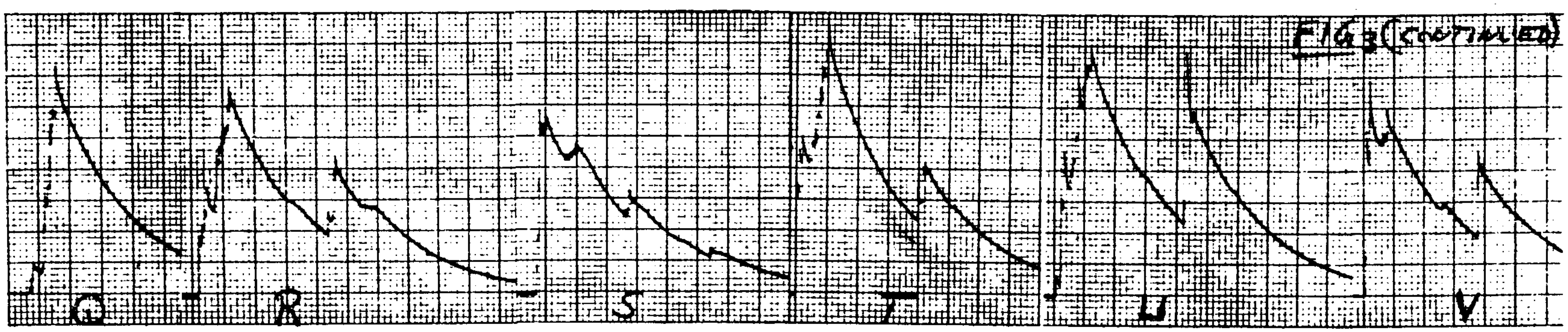


FIG 3 (CONTINUED)
Q
R
S
T
U
V

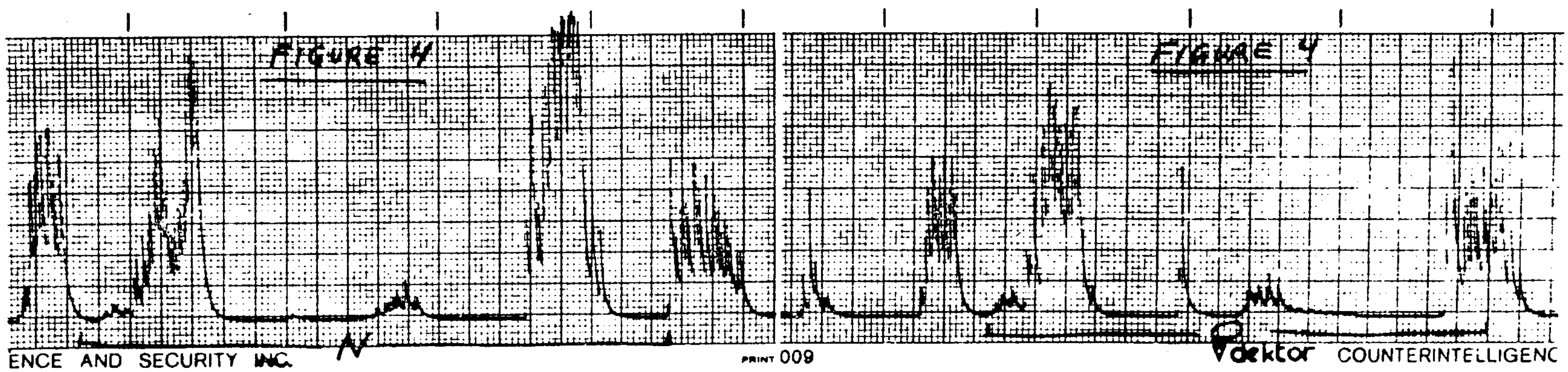


FIGURE 4
FIGURE 4
N
ENCE AND SECURITY INC.
PRINT 009
Dektor COUNTERINTELLIGENC

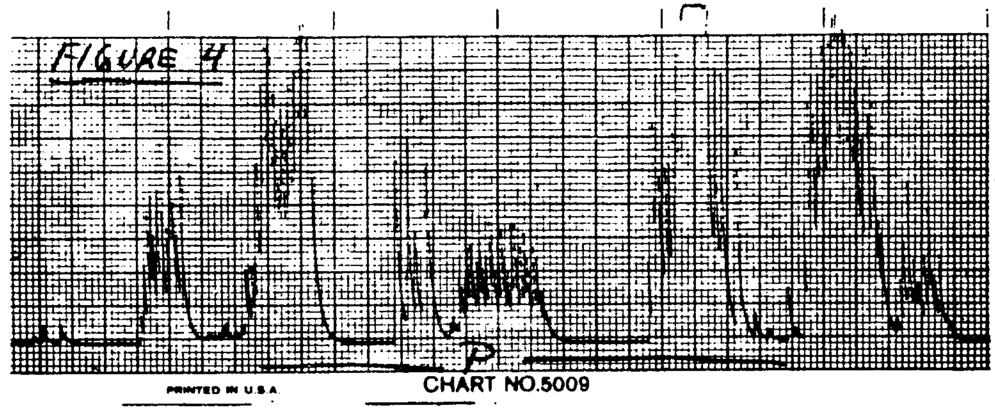


FIGURE 4
PRINTED IN U.S.A.
CHART NO.5009

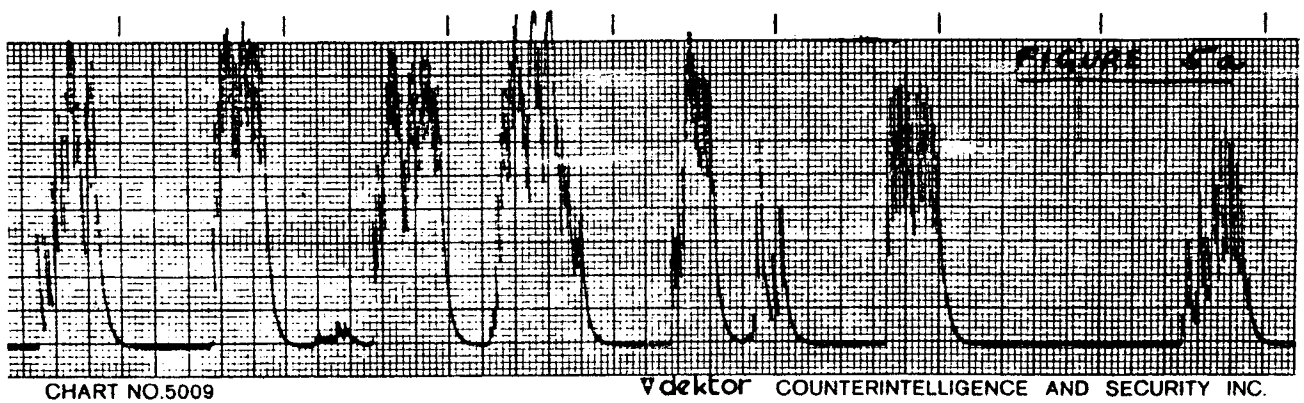


FIGURE 5a
CHART NO.5009
Vdektor COUNTERINTELLIGENCE AND SECURITY INC.

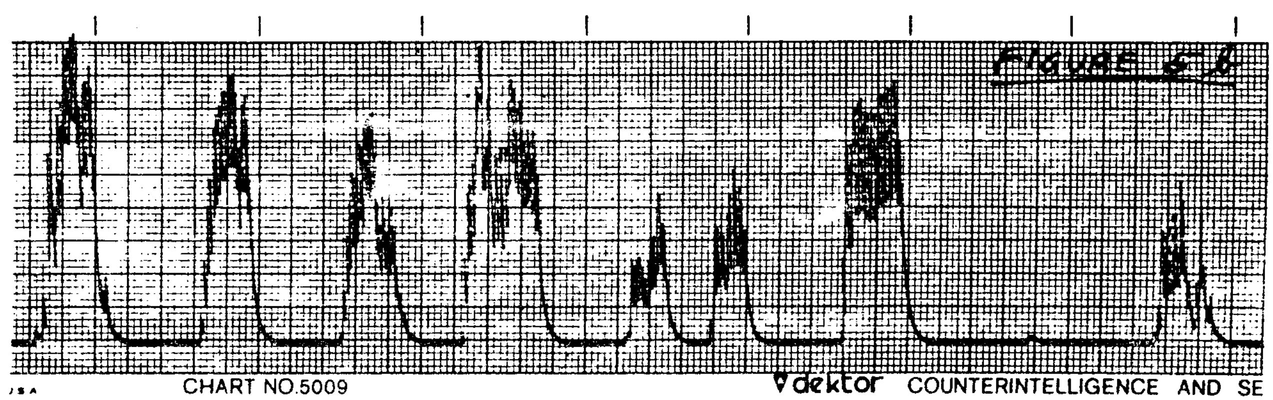


FIGURE 5b
CHART NO.5009
Vdektor COUNTERINTELLIGENCE AND SE