TECHNICAL NOTE

Voice Analysis to Predict the Psychological or Physical State of a Speaker

ROBERT RUIZ, M.S., CLAUDE LEGROS, M.S., Ph.D., and
ANTONIO GUELL, M.D.

RUIZ R, LEGROS C, GUELL A. *Voice analysis to predict the psychological or physical state of a speaker.* Aviat. Space Environ. Med. 1990; 61:266–71.

A vocal message, apart from its semantic content, carries information on the psychological and physiological condition of the speaker. Physical fatigue and especially psychological stress are the pathological elements of the condition. The accepted term for the cause of these effects is the "workload." This article describes the main research carried out since the 1940's to measure the acoustic modifications of the voice brought about by a workload. It concludes by a critical analysis of the studies and a short description of the perspectives for research. Their results mainly concern astronauts and pilots involved in specific high-stress tasks and possible users of voice recognition systems. All the studies show an excellent approach to this field of research but deserve to be widened, deepened, and made more accurate to enable estimating the nature or level of reaction to a workload.

WORKLOAD, usually not defined, can bring about psychological and physiological disturbances that are the measurable signs of reaction by the individual, who changes from a state of rest or relaxation to one of physical fatigue or psychological or emotional agitation (stress).

It appears that these two states can be differentiated by acoustic analysis of the voice. Although the main role of the voice is to communicate, it is also probable that the voice is an indicator of the psychological and physiological state of the speaker.

The identification of vocal acoustic indicators allowing evaluation of a person's reaction to a workload can enable remote diagnosis of the psychophysiological state of a speaker without the speaker's awareness.

This article summarizes, analyzes, and comments on the main results of research to date. The need for the present review has arisen from the relative diversity and heterogeneity of the studies carried out. Moreover, this original research subject responds to a need which, from the 1940's, has still not been satisfied and has continued to grow with the development of technology, especially in aviation and space. It has become necessary to monitor the psychological and physiological state of the people involved in order to prevent failures which could cause heavy human and material losses. Also, the ever-increasing use of command and voice recognition systems implies that they must operate in spite of any modifications of the acoustic characteristics of the voice.

One-Dimensional Acoustic Characteristics

Frequency characteristics: The research indicates that the fundamental frequency F_0 and the frequency of the first formants seem to be the frequential acoustic characteristics of the voice which undergo modifications when the subject is under the influence of a workload.

Except for one case (5) the studies all conclude that the fundamental frequency rises with the level of the subject's emotional stress or the complexity of the task performed (2,3). For example, Griffin (3) showed that, whereas the mean resting fundamental frequency is 106.69 Hz (σ = 12.66 Hz), it rises to 108.58 Hz (σ = 13.57 Hz) during performance of a first psychomotor task, then to 115.4 Hz (σ = 14.05 Hz) with a second task, and finally to 118.91 Hz (σ = 14.95 Hz) for a workload combining both. The analysis of air-to-ground radio communications also confirms the rise of F_0 (Table I). Whatever the type of workload, all the research comes to the same conclusion: the mean fundamental frequency of the voice in a stressful situation is higher than in a situation of rest.

It seems that the frequencies of the first two formants, and even the third, increase when a person is subjected to a psychomotor task (14) or simulates emotions (anger, fear, sorrow) (24). A Soviet study (18) indicates that the variations of the first formant frequency allow emotional stress to be evaluated: the number and amplitude of first formant maximum passages in five third-octave filters is greater than in normal speech.

From the Laboratoire d'Acoustique, de Métrologie et d'Instrumentation, and the Centre National d'Etudes Spatiales, Toulouse, France.

This manuscript was received for review in May 1989. The revised manuscript was accepted for publication in August 1989.

TABLE I. MEDIAN FUNDAMENTAL FREQUENCY AND RANGE (10%, 90%) OF FUNDAMENTAL FREQUENCY FOR EACH OF FOUR TALKERS SPEAKING DURING TWO SITUATIONS.*

Talkers	Median F_0	F_0 Range
Civilian pilot		
a) before serious flight difficulty	208 Hz	168–272 Hz
b) during flight difficulty prior to crash	432 Hz	288–492 Hz
Civilian control tower operator		
a) before pilot's report of difficulty	210 Hz	168–228 Hz
b) after pilot's distress call	284 Hz	260–308 Hz
Military control tower operator		
a) before loss of radio contact with pilot	112 Hz	100–140 Hz
b) after loss of radio contact with pilot	151 Hz	124–176 Hz
Radio announcer		
a) before crash of "Hindenburg"	166 Hz	124–196 Hz
b) after crash of "Hindenburg"	196 Hz	152–260 Hz

* From Williams and Stevens (23).

Energy levels: The mean sound pressure level decreases with time by about 3 dB compared to the resting state when the speakers have been in total darkness for 10 min. It increases by about 1 dB when they are subjected to weak electric shocks on the skin during speech and then decreases by about 2 dB when the shocks are removed (16).

The average value of peak sound pressure level vs. time increases with the complexity of the workload (psychomotor task or dichotic listening): it is 49.38 dB ($\sigma = 8.08$ dB) for the voice of resting subjects, 49.43 dB ($\sigma = 7.45$ dB) for the first experiment (psychomotor task), 55.82 dB ($\sigma = 8.47$ dB) for the second (dichotic listening), and 57.12 dB ($\sigma = 8.68$ dB) for the simultaneous performance of both (3). The results of another experiment with a tracking task confirm the rise of average amplitude (2).

The average value of the peak sound pressure level measured for the utterance of one or more specific phonemes, first when the speaker is resting, then when carrying out mental exercises, does not always vary in the same way (5).

Rates and durations: The speaking rate in words/min or in syllables/s decreases when the speaker simulates emotions (anger, fear, grief); the total speaking time, however, increases (25) (Tables II, III).

Word duration increases under the action of a psychomotor task, but then decreases as the task becomes more complex. The results of the experiment are: resting 384.81 ms ($\sigma = 57.12$ ms); psychomotor task 388.01 ms ($\sigma = 55.20$ ms); dichotic listening 342.95 ms ($\sigma = 47.87$ ms); combination of both 338.80 ms ($\sigma = 14.95$ ms). The decrease observed in word duration is justified by the increase in speaking rate for the workload studied (3).

TABLE II. MEAN SPEAKING RATE (syllables/s).*

	Neutral	Anger	Fear	Sorrow
Voice A	4.03	4.26	3.92	1.84
Voice B	4.89	4.32	3.90	2.03
Voice C	4.02	3.88	3.57	1.86
Mean	4.31	4.15	3.80	1.91

* From Williams and Stevens (24).

TABLE III. TOTAL SPEAKING TIME DURATION (seconds) AND SPEAKING RATE (words/min).*

	Total Speaking Time	Rate
Indifference	7.74	209
Fear	8.03	202
Anger	8.51	190
Grief	12.57	129
Contempt	14.03	116

* From Fairbanks and Hoaglin, 1941. See (25).

The measurement of various geometrical distances along the time axis of a wide-band sonagram (i.e., the time period of vocal cords) enabled the authors (9) to propose a method for determining the state of emotional stress from air-to-ground radio communication recordings of pilots having experienced flight incidents with fatal consequences. The indicator calculated for this purpose was the Vibration Space Shift Rate (VSSR). VSSR in percent = [(SVS-EVS)/SVS] • 100, where SVS (Standard Vibration Space) represents the greatest geometrical distance in micrometers between the vertical deflections of a vowel on the sonagram of the pilot's voice during a normal flight phase; and EVS (Emergency Vibration Space) is the same measurement for the emergency phase of the same flight.

The VSSR is therefore a simple indicator of the behaviour of the fundamental frequency in a stressful situation. The statistical analysis of the results shows that 70% of voices during the normal flight phase have a VSSR lower than 17.24%; 70% of voices during the emergency phase have a VSSR between 17.25% and 42.74%, and 70% of voices in the moments before ejection have a VSSR greater than 42.75% (Fig. 1). Comparison of VSSR measurements made at different times during radio communication and measurements of the fundamental frequency at the same times shows that the two magnitudes vary in the same way. The increase in emotional stress is associated with an increase in both the VSSR and the fundamental frequency.

Two-Dimensional Acoustic Characteristics

Time contour of the fundamental frequency F_0: Several authors have observed that the time contour of the fundamental frequency undergoes modifications when the speaker is in a stressful situation. The contours are drawn from periodic measurement of the coarse or slight variations of the fundamental frequency.

Comparison of resting voice contours with those observed under stress is rather subjective (i.e., visual). When the speaker performs a mental arithmetic task, the fluctuations of the fundamental frequency become smaller with time. The authors link the pattern of the

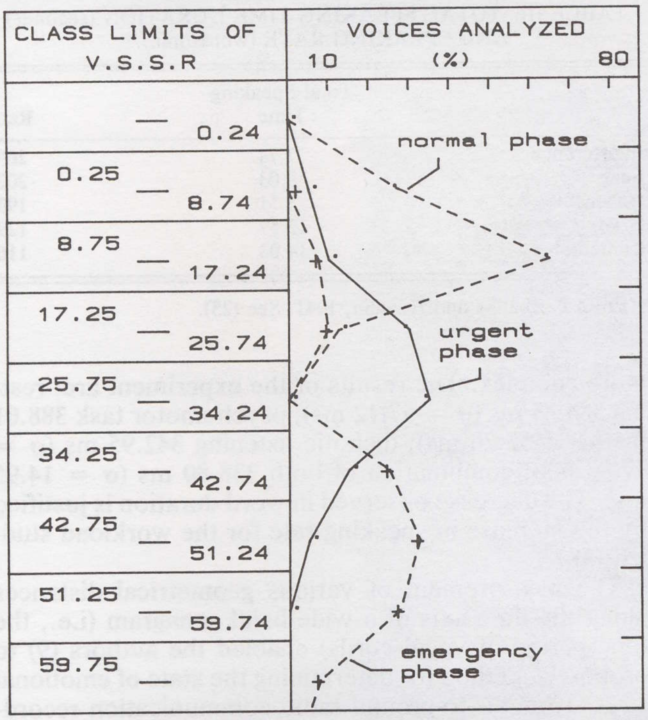

Fig. 1. Distribution of Vibration Space Shift Rate values for normal, urgent, and emergency flight phases. From Kuroda (9).

time contour with the monotony of the voice (5). However, the time contours of the fundamental frequency become more irregular and discontinuous in radio communications between pilots and ground staff when technical incidents occur. Sudden changes, or jumps, in the frequency of F_0 and rapid fluctuations from one syllable to the next can be observed (23). Experiments on vocal simulation of various emotions lead to the same conclusions. Vocal expressions of anger, fear, or grief show F_0 contours which are more irregular than during a neutral emotional situation (24,25). The authors consider that these irregularities represent a loss of accurate control of the muscles involved in articulation and those of the larynx, and indicate modifications in breathing. This disorganisation of motor activity implies lack of coordination between the movements of the tongue, lips, and jaw and the frequency control of the larynx muscles.

The Psychological Stress Evaluator (PSE) supplies the experimenter with the frequency/time pattern of the fundamental frequency F_0. A slow frequency modulation of the fundamental frequency of the vocal signal (8–14 Hz) is seen during speech in an emotionally neutral situation. This is reportedly caused by the natural physiological trembling of the muscles around the vocal cords. When the speaker is subjected to the action of a mental or psychomotor workload, the modulation of F_0 deteriorates with the degrees of reaction of the individual: muscle contraction tends to cancel out the modulation. The interpretation of the diagrams supplied by PSE is essentially subjective (i.e. visual) (21). The task was however performed by a computer which was programmed to detect three types of characteristic patterns (high stress, medium stress, and low stress) and which, by comparison with the diagram obtained with the PSE, gave a score from 0–10 (Fig. 2); the highest score corresponds to the highest stress situation (17).

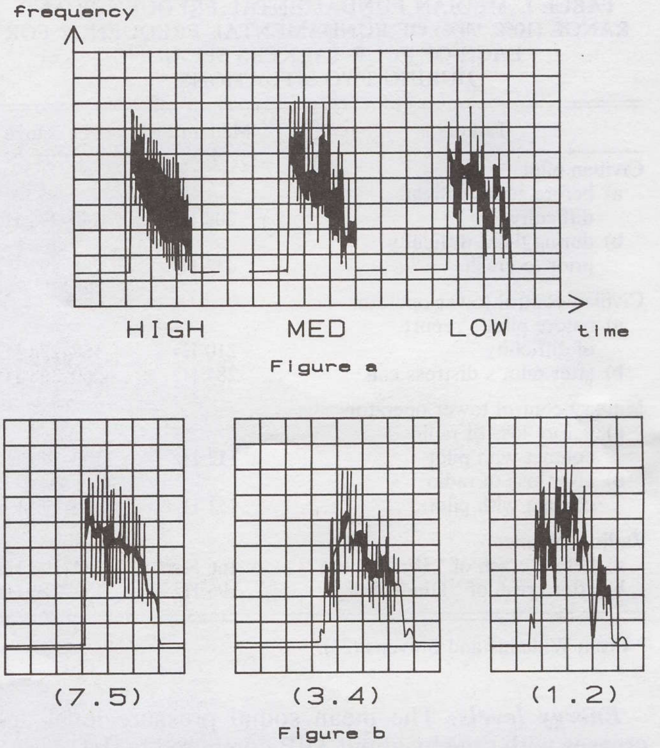

Fig. 2. (A) Model patterns of three different types of PSE diagrams corresponding respectively to a high, medium and low stress vocal utterance. These diagrams are stored in a computer. (B) Attribution of a coefficient (0–10) by a computer comparing every new diagram with the three above model patterns. From Schiflett and Loikith (17).

Most of the experiments described in the literature concern the evaluation of emotional stress brought about by lying (1,6). They attempt to justify or reprove the use of the PSE as a lie detector—the function for which the PSE is sold. The PSE has also been tested for the observation of emotional stress during radio broadcasts (21), the influence of psychomotor workloads (1,17), and the influence of the operational tasks performed by the astronauts in Skylab (15). Analyses of the research concerning the use of PSE in lie detection suggest that it cannot be used for this because it is not error-free. The PSE analyses are not beyond chance levels and the device seems sensitive to conscious control of emotions (1,6). However, for workloads which bring about a stronger individual reaction, it seems that significant statistical relationships exist between the PSE diagrams and the state of stress. Yet, the voice analysis technique used by the PSE is not predictable enough to guarantee its use in the evaluation of the level of psychological stress, especially with a low "quantity of stress."

Frequency pattern of the voice energy: Through one-third-octave spectral analysis of isolated words, it seems possible to correlate emotional stress with an indicator, M, over the frequency range of the first formant (300–1200 Hz) for non-simulated laboratory workloads such as analysis of radio messages with $U \geq W$ (18,20).

$$M = \frac{\sum_{1}^{U} (f_i \cdot P_i)}{\sum_{1}^{U} P_i} \cdot \log C \cdot \sum_{1}^{W} \left(\frac{P_i}{P}\right)$$

where, for the whole duration of the voice signal, we have: P_i: mean power at the output of the i^{th} one-third-octave filter; P: mean power at the output of a linear filter; f_i: center frequency of the i^{th} one-third-octave filter; C = 1000: proportionality coefficient; U and W are the ordinal numbers of the one-third-octave filters, W = U when U = 1 or 2, and W = 2 when U > 2.

According to the authors, an increase in emotional stress of a subject causes a rise in the numerical value of M. This method enables them, in 85% of cases, to estimate the degree of emotional stress. Moreover, they observe a correlation with the direction of variation of $E = (M/M_{max})$ and of the heart rate for successive work phases of cosmonauts in flight (Fig. 3). The same Russian authors also proposed a method to estimate the level and the nature of emotional stress of actors vocally simulating fear, anxiety, joy, and delight. In order to do this, they evaluated the following magnitudes:

$$\alpha_i = \int_{0}^{T} U_i(t) \cdot dt$$

$$\gamma = \frac{\alpha_1}{\alpha_2} + \frac{\alpha_1}{\alpha_3} + \frac{\alpha_2}{\alpha_3}$$

$$\beta_1 = \frac{\min(\alpha_2,\alpha_3) + \alpha_1}{\max(\alpha_2,\alpha_3)}$$

$$\beta_2 = \frac{\alpha_4 + \alpha_6}{\min(\alpha_2,\alpha_3) + \alpha_1}$$

$$\beta_3 = \frac{\alpha_1}{\min(\alpha_4,\alpha_6)}$$

$$\beta_4 = \frac{\alpha_5}{\min(\alpha_2,\alpha_3)}$$

where $U_i(t)$ is the envelope of the vocal signal at the output of octave filter i from i = 1 (125 Hz) to i = 6 (4000 Hz) and where T is the duration of the vocal signal.

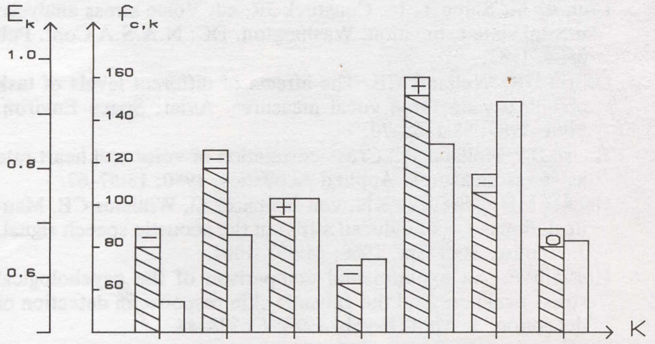

Fig. 3. Relationship between heart rate ($f_{c,k}$, clear bar) and the index of the degree of emotional stress (E_K, hatched bar) for different phases (K) of the Voskhod-2 space flight. Psychological signs of emotion (upper part of each hatched bar) are negative (−), positive (+), and uncertain (○). K = 1: pre-launch situation, K = 2: minute readiness, K = 3: active stage, K = 4: preparation for extra vehicular activity (EVA), K = 5: exit into outer space, K = 6: EVA, K = 7: in the chamber. From Simonov and Frolov (20).

When the subject is resting, the energy contribution of the lower frequencies to the vocal signal is always greater than when the speaker is in a situation of emotional stress (γ is highest at rest and decreases as stress rises). The decrease enables the authors to evaluate the degree of stress. Comparison of the products $\beta_1 \cdot \beta_3$ and $\beta_2 \cdot \beta_4$ indicates the nature of stress. If $\beta_1 \cdot \beta_3 > \beta_2 \cdot \beta_4$, the emotion involved is "negative" (i.e., fear, anxiety); if $\beta_1 \cdot \beta_3 < \beta_2 \cdot \beta_4$, the emotion is "positive" (i.e., joy, delight). According to the authors, the method is efficient in 90% of the emotion simulations (Fig. 3,4).

Spectral analysis of the voices of actors simulating various emotions was also carried out by other authors. They studied the spectrum over octave bands from 125–4,000 Hz (24). The levels from each filter were averaged over several seconds and expressed as a fraction of the level of the 250-Hz filter. The relative sound level in the 125-Hz band gives a rough indication of the energy-frequency behaviour of the fundamental frequency F_0, with the type of vocal emotion being simulated by the actors. Thus, emotions for which it had been observed that the fundamental frequency was high (greater than 175 Hz in the present case) give rise to a relative lower level in the 125-Hz band compared to the rest. This spectral analysis confirms the rise in frequency of F_0 during the vocal expression of various emotions. It also allows the experimenter to compare the energy changes at frequencies greater than or equal to 500 Hz. This shows that speech in a state of anger or fear has a higher energy above 1000 Hz than "neutral." The opposite is noted for sorrow.

Three-Dimensional Acoustic Characteristics

The comparative observation of the general shape and of particular characteristics of an individual's voice sonagram at rest and under stress show these characteristics are of the qualitative type. The influence of workload brings about modifications of the amount of energy at frequencies above 2,000 Hz (degree of darkening of the sonagrams). Changes in shape of the sonagrams can also be observed (e.g., attack, regularity of successive pulses, accuracy of articulation, slurring of

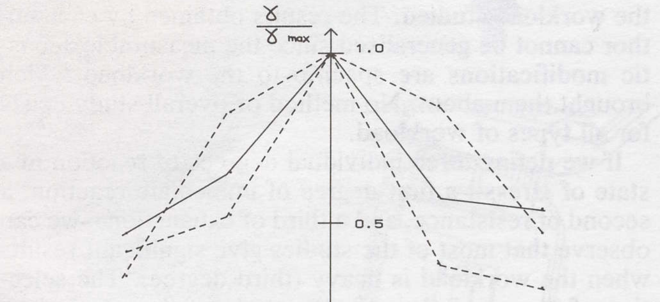

Fig. 4. Graphic characteristics of the method of differentiation of positive and negative emotions by calculation of γ/γ_{max}, as in the text from Simonov and Frolov (18).

vocal expression between syllables, syntactic omissions). None of these characteristics is easy to quantify. They vary considerably from one speaker to another. Nevertheless, comparison of sonagrams does enable them to be detected.

Conclusion

To date, the aim of the studies has been the identification and measure of acoustic characteristics of the voice which, through their modification, indicate the influence of various workloads. The results obtained have not been sufficient to enable an operator in possession of vocal recordings of subjects presumed to be in a state of stress to discover the type and degree of the disturbance.

The statistical analyses of the experiments led the authors to note that the numerical differences between the characteristics measured in the resting state and in a state of stress are significant. The modifications of the acoustic indicators brought about by the influence of the workload are observed even though the variations are only of a few Hertz, decibels, or milliseconds. Yet if we take one numerical value of a one-dimensional characteristic for the resting state and the stressed state it is not possible formally to identify the nature or the degree of stress brought about by the workload. This is either because the variability of the indicators between speakers is too great or their modifications are too slight. This suggests that the acoustic characteristics chosen as indicators are not sufficiently sensitive to the effects on the voice of the psycho-physiological disturbances brought about by the workload.

The two- and three-dimensional acoustic characteristics are really indicators showing the existence of modifications, rather than effective estimators of the reaction of an individual to a workload. Essentially, they allow the authors to observe, in an approximate way, that a correlation does exist between the stress and voice.

The research to date is rather incomplete. There are not enough studies on modifications of voice characteristics in a state of physical fatigue. The experimental choice of laboratory workloads is rather subjective. It is not possible to know if laboratory results with artificial or simulated workloads are useful for the study of actual workloads. The acoustic indicators are closely linked to the workload studied. The results obtained by each author cannot be generalised since the measurable acoustic modifications are specific to the workload which brought them about. No method or overall study exists for all types of workload.

If we define three individual degrees of reaction in a state of stress—a first degree of immediate reaction, a second of resistance, and a third of exhaustion—we can observe that most of the studies give significant results when the workload is heavy (third degree). The selection of the vocabulary for the study of the vocal characteristics of stress has not been the subject of any special study. The choice is more a matter of facilitating experimentation (e.g., pronouncing numbers, yes/no tests) than on justifying the statistical analysis of the acoustic modifications to the phonetic elements of the language. Such an analysis would determine the elements of vocabulary or of the acoustico-phonetic framework which are the most sensitive to the influence of the workload for the acoustic indicators under study. To our knowledge, no reports have appeared concerning the lexico-grammatical and semantic influence of a workload. The mean number of subjects who executed the experimental tests is 15. This is not large enough to give a correct statistical representation of workload influence. The overall results are tangible signs of the existence of acoustic modifications of the vocal signal. They do not yet allow an actual quantification of the vocal manifestations of psycho-physiological disturbances brought about by a workload.

Some acoustic signs of the vocal signal have been determined to indicate the individual reaction. However, many uncertainties remain due to the difficulty of the research, so prospects are numerous and varied. The investigations must give an exact definition of the workloads. The acoustical observations are closely linked to the workload studied. There is a need to evaluate ‘‘the stress power’’ of every kind of workload by comparing their influence on the voice. The research perspectives consist of verifying the following properties of acoustic characteristics: low intra-variability for the speaker at rest and high intra-variability under the influence of a workload; inter-speaker variability as low as possible; and monotonous change of acoustic characteristic with increasing workload. They consist of using a non-specific vocabulary and making possible the ‘‘continuous’’ measurement of the acoustic characteristics during performance of the workload.

The ultimate aim of the research is to develop a forecast method; the pathological psycho-physiological condition of an individual carrying out a workload should be predictable from a mere analysis of the voice.

ACKNOWLEDGMENTS

This work was supported by the Centre National d'Etudes Spatiales.

REFERENCES

1. Brenner M, Branscomb HH, Schwartz GE. Psychological stress evaluator—two tests of a vocal measure. Psychophysiology 1979; 16:351–7.
2. Brenner M, Shipp T. In: Constuck JR, ed. Voice stress analysis: mental state estimation. Washington, DC: N.A.S.A Conf. Pub 2504, 1987.
3. Griffin GR, Williams CE. The effects of different levels of task complexity on three vocal measures. Aviat. Space Environ. Med. 1987; 58:1165–70.
4. Hayre HS, Holland JC. Cross-correlation of voice and heart rate as stress measures. Applied Acoustics, 1980; 13:57–62.
5. Hecker MHL, Stevens KN, von Bismarck G, Williams CE. Manifestations of task-induced stress in the acoustic speech signal. J. Acoust. Soc. Am. 1968; 44:993–1001.
6. Horvath F. An experimental comparison of the psychological stress evaluator and the galvanic skin response in detection of deception. J. Appl. Psych. 1978; 63:338–44.
7. Kachatur'yants L, Grimak L. Cosmonaut's emotional stress in space flight. Washington, DC: N.A.S.A TT F-14. 654, 1972.
8. Kozlovsky AP, Kovalenko AF. Psychological (emotional) stress in pilots awaiting ejection and its role in task performance. Washington, DC: N.A.S.A contractor report U.S.S.R Space Life Sciences Digest, April 1986, p. 225.
9. Kuroda I, Fujiara O, Okamura N, Utsuki N. Method for determining pilot stress through analysis of voice communication. Aviat. Space Environ. Med. 1976; 47:528–33.

10. Kuznetsov V, Lapayev E. Voices in orbit. Washington, DC: N.A.S.A. TT F-16499.
11. Legros C, Ruiz R. Etude de l'influence d'une charge de travail sur les caractéristiques acoustiques de la voix. 1. bibliographie. Toulouse, France: Rapport interne du Centre National d'Etudes Spatiales, Septembre 1988.
12. Lieberman P. Perturbations in vocal pitch. J. Acoust. Soc. Am. 1961; 33:597–603.
13. Lieberman P, Michaels, SB. Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech. J. Acoust. Soc. Am. 1962; 34:922–7.
14. Mosko JD, Stevens KN, Griffin GR. Interactive voice technology: variations in the vocal utterances of speakers performing a stress-inducing task. Pensacola, FL: Naval Aerospace Medical Research Laboratory, 1983, N.A.M.R.L. 1300.
15. Older HJ, Jenney LL. Psychological stress measurement through voice output analysis. Washington, DC: N.A.S.A. CR 141723, 1975, N 75-19960.
16. Rubenstein L. Electro-acoustical measurement of vocal responses to limited stress. Behav. Res. Ther. 1966; 4:135–8.
17. Schiflett SG, Loikith GJ. Voice stress analysis as a measure of operator workload. Patuxent River, MD: Naval Air Test Center, 1980, TM 79-3 SY.
18. Simonov PV, Frolov MV. Utilization of human voice for estimation of man's emotional stress and state of attention. Aerospace Med. 1973; 44:256–8.
19. Simonov PV, Frolov MV, Taubkin VL. Use of the invariant method of speech analysis to discern the emotional state of announcers. Aviat. Space Environ. Med. 1975; 46:1014–6.
20. Simonov PV, Frolov MV. Analysis of the human voice as a method of controlling emotional state: achievements and goals. Aviat. Space Environ. Med. 1977; 48:23–5.
21. Smith GA. Voice analysis for the measurement of anxiety. Br. J. Med. Psychol. 1977; 50:367–73.
22. Streeter LA, Krauss RM, Geller V, Olson C., Apple W. Pitch changes during attempted deception. J. Personality Social Psychol. 1977; 35:345–50.
23. Williams CE, Stevens KN. On determining the emotional state of pilots during flight: an exploratory study. Aerospace Med. 1969; 40:1369–72.
24. Williams CE, Stevens KN. Emotions and speech: some acoustical correlates. J. Acoust. Soc. Am. 1972; 52:1238–50.
25. Williams CE, Stevens KN. Vocal correlates of emotional states. In: Darby JK, Jr, ed. Speech evaluation in psychiatry. New York: Grune & Stratton, Inc., 1981.