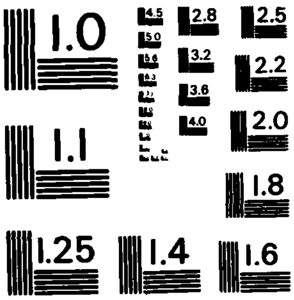
AD A132 577 STRESS ASSESSMENT THROUGH VOICE ANALYSIS(U) TECHNOLOGY INC SAN ANTONIO TX LIFE SCIENCES DIV N C CHAMBERS ET AL. SEP 83 AFHRL-TP-83-47 1/1

UNCLASSIFIED F33615-80-C-0018 F/G 17/2 NL

END
DATE FILMED
9 83
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AFHRL-TP-83-47

AIR FORCE

HUMAN RESOURCES

STRESS ASSESSMENT THROUGH VOICE ANALYSIS

By

N.C. Chambers
J.C. Brakefield
D.I. Yahiel
D.D. Fulgham

Technology Incorporated
Life Sciences Division
300 Breesport
San Antonio, Texas 78216

MANPOWER AND PERSONNEL DIVISION
Brooks Air Force Base, Texas 78235

September 1983
Final Technical Paper

Approved for public release; distribution unlimited.

LABORATORY

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235

NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The Public Affairs Office has reviewed this paper, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This paper has been reviewed and is approved for publication.

NANCY GUINN, Technical Director
Manpower and Personnel Division

J.P. AMOR, Lt Col, USAF
Chief, Manpower and Personnel Division

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (*When Data Entered*)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFHRL-TP-83-47	2. GOVT ACCESSION NO. AD-A132577	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (*and Subtitle*) STRESS ASSESSMENT THROUGH VOICE ANALYSIS		5. TYPE OF REPORT & PERIOD COVERED Final
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(*s*) N.C. Chambers, J.C. Brakefield, D.I. Yahiel, D.D. Fulgham		8. CONTRACT OR GRANT NUMBER(*s*) F33615-80-C-0018
9. PERFORMING ORGANIZATION NAME AND ADDRESS Technology Incorporated, Life Sciences Division 300 Breesport San Antonio, Texas 78216		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 62703F 77191816
11. CONTROLLING OFFICE NAME AND ADDRESS HQ Air Force Human Resources Laboratory (AFSC) Brooks Air Force Base, Texas 78235		12. REPORT DATE September 1983
		13. NUMBER OF PAGES 48
14. MONITORING AGENCY NAME & ADDRESS (*if different from Controlling Office*) Manpower and Personnel Division Air Force Human Resources Laboratory Brooks Air Force Base, Texas 78235		15. SECURITY CLASS (*of this report*) Unclassified
		15.*a.* DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (*of this Report*)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (*of this abstract entered in Block 20, if different from Report*)

18. SUPPLEMENTARY NOTES

This technical paper reports the results of a feasibility study and is being disseminated for documentation purposes only. It has not been edited by AFHRL. The opinions expressed are those of the contractor and do not necessarily represent those of the United States Air Force.

19. KEY WORDS (*Continue on reverse side if necessary and identify by block number*)

recognition
speech analysis
stress
voice

20. ABSTRACT (*Continue on reverse side if necessary and identify by block number*)

An attempt was made to assess emotional stress through voice analysis. The involuntary vocal musculature microtremor, which has been reported in some literature as correlated with stress, was chosen as the variable for analysis. An autocorrelation technique was developed using a CD&A array processor installed in a DEC PDP 11/34 computer. Although the technique isolated FM, the microtremor was not identified, primarily due to excessive noise in the 5-15 Hz frequency band of interest.

AFHRL Technical Paper 83-47 September 1983

STRESS ASSESSMENT THROUGH VOICE ANALYSIS

By

N.C. Chambers
J.C. Brakefield
D.I. Yahiel
D.D. Fulgham

Technology Incorporated
Life Sciences Division
300 Breesport
San Antonio, Texas 78216

MANPOWER AND PERSONNEL DIVISION
Brooks Air Force Base, Texas 78235

This working paper is being disseminated for documentation purposes only. It has not been edited by AFHRL. The opinions expressed herein are those of the contractor and do not necessarily represent those of the United States Air Force.

Preface

This study was accomplished under work unit 77191819, Voice Spectral Analysis as a Measure of Stress in Air Combat. This work unit is part of the Laboratory program on Personnel Qualifications which supports the research thrust area, Force Acquisition and Distribution Systems. This particular work was a feasibility study exploring a potentially useful new technology. The primary reason for documenting this effort is to provide a medium for lessons learned. The lack of success of this research might be attributable to two possible causes: (1) an inappropriate R&D approach, or (2) the non-existence of a consistent and measurable change in a person's voice under conditions of stress.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

In human languages, the change of the fundamental frequency (F_0) over time contributes both linguistic and paralinguistic information to the total message articulated. In many languages, among which Chinese is the best known example, controlled changes in fundamental frequency are phonemic and used for linguistic purposes. The same vowels and consonants signify different words when different F_0 patterns are employed; therefore, the lexical meaning of the word depends upon the type of F_0 contour it contains. In English, F_0 changes are said to be nonphonemic, since the lexical meaning of a word cannot be altered by a change in the F_0 contour over the course of the word. Nonetheless, the type of information conveyed by the fundamental frequency in English is far from unimportant. In conjunction with the cues provided by intensity and duration, the changes of F_0 in English conveys to the listener whether a statement is being made or a question is being asked; which syllable in a word is being stressed or which word in a sentence is being emphasized; and, whether an utterance reflects surprise, dismay, assuredness or shock on the part of the speaker. It is the acoustic correlates of these prosodic and suprasegmental features of speech that give some indication of the emotional state of the individual. Williams and Stevens (1972) suggest that the F_0 of the speech signal versus time appears to be the clearest indicator of the emotional state of the speaker.

Numerous researchers (Hecker, 1971; Lynch, 1934; Fairbanks and Provonost, 1939; Fairbanks and Hoaglin, 1941; Lieberman, 1961; Lieberman and Michaels, 1962; Fonagy and Magdies, 1963; Uldall, 1960) have investigated the relationship between speech and the artificial simulation of emotions. Only a few studies (Skinner, 1935; Huttar, 1968) have used normal speech in their experiments. All of the above studies, both real and simulated, focused on connected speech. Hollien, et al. (1973) used sustained phonation rather than connected speech in an attempt to determine whether the reported changes in cycle-to-cycle variation were due to (1) involuntary, inherent phonatory variation (jitter), or (2) voluntary and/or learned inflectional speech patterns. Results suggested that the degree of laryngeal jitter increased as a function of the phonated frequency. The jitter factors of 0.5-1.0 were considered average limits for sustained phonation by normal males. A study by Beckett (1969), also based on normal male subjects sustained phonations, investigated the relationship of pitch perturbation to three levels of vocal constriction. Results indicated that the measure of pitch perturbation was a function of subjective vocal constriction. Utilizing synthetic vowels, Rozsypal and Miller (1979) applied a multi-dimensional scaling technique in the analysis of jitter and shimmer. Their experiment indicated that (1) some jitter is necessary for sustained vowels to be perceived as natural, (2) the vowel sound determines the optimal amount of jitter, and (3) the shimmer effect is equal for all vowels and less pronounced than that of jitter.

At present, an area receiving much attention in voice analysis of stress is the muscle microtremor phenomena. The microtremor, also referred to as involuntary voice tremor, involuntary frequency modulations (FM), speech tremor, and pitch perturbations is stirring professional interest in the fields of acoustic phonetics, aviation, law enforcement, and psychiatry.

It has been hypothesized that the voice microtremor is related to the phenomenon of physiological tremor which was discovered many years ago to be a normal accompaniment of a voluntary muscle activity. Lippold (1971) noted that the normal contraction of a voluntary muscle is accompanied by tremors of the muscles which take the form of minute oscillations which are diminished with excitation of the muscle source. The frequency characteristics of these oscillations, which occur between 8 and 12 cycles per second, were isolated by Halliday and Redfearn through the use of Fourier analysis (Edson, 1976). The application of these research findings were not utilized in voice stress analysis until the development of the Psychological Stress Evaluator (PSE), a deception detection instrument, by Bell, Ford and McQuistin (1972).

The popularity of this voice analyzing equipment (Dektor, 1971) with law enforcement agencies, psychiatric clinics, private investigators, etc., is based upon the manufacturer's claim that the PSE discerns a physiological tremor of the voice mechanism which is present in a relaxed emotional state, and disappears with psychological stress. Furthermore, the involuntary vocal tract tremor, which is superimposed upon the F_0, is under the control of the central nervous system until it is suppressed by the autonomic nervous system which gains dominance in a stress situation.

Some reports question the validity of the PSE and indicate that it only works under acute or high stress conditions (McGlone, et al., 1974; Papcun, 1974; Lambert, 1974). However, studies by Smith (1977) and Eden and Inbar (1975, 1976, 1978) support the PSE as an instrument capable of measuring anxiety. Smith's study indicated that "stress blocking" of the voice patterns appeared where it was expected to appear, however, more accurate and objective scoring systems were needed. Inbar and Eden's three-part study (1975, 1976, 1978) confirms the statements by the PSE proponents that the central nervous system is the source of the vocal tract tremor. In the first part of the study (1975) electromyogram (EMG) correlates of the PSE were sought through the utilization of two methods: (1) transcutaneous stimulation of the vocal tract muscles by external surface electrodes to verify the ability of muscle tension changes to generate correlated voice tremor, and (2) a throat microphone to detect tremor type vibrations in the pitch waveform. Positive results were obtained from the first method. The second method revealed tremor vibrations detected in the first formant of regular speech were also found in the pitch waveform. In the second part of the study (1976), the hypothesis that the frequency changes in speech are controlled by the central nervous system was investigated. In this experiment, surface EMG recordings were used to estimate changes in the tension of muscles in the vocal area. Results indicated that voice tremor can be produced in two ways: (1) by mechanical subresonances in the vocal cords or vocal tract, and (2) by signals generated by the central nervous system. Cross-correlation results indicated that the voice tremor is produced by the central nervous system. The evidence to support this conclusion is that the oscillations were random in nature, and always preceded by voice tremor by approximately the same amount of time for a particular vowel. Because of this finding, EMG tremor could not originate from muscle spindle afferent reflex signals activated by mechanical sources, but only by central nervous system activation. In the final segment of Inbar and Eden's study (1978), two theoretical aspects were tested: (1) the influence of pitch period variations on frequency changes resulting from the resonant characteristics of the vocal tract; and (2) the vocal system's physiological parameters which are potentially able to govern involuntary frequency changes. Results indicated that activity of both the vocal cords and the vocal tract can produce frequency variation in the human voice.

In summary, in preparation for this effort, an extensive literature review documented studies relating to the varying emotional states of the voice and the acoustical correlates of the prosodic features of the voice: intensity, fundamental frequency variation (microtremor) and spectrum patterns of intonation. It was decided to concentrate on the vocal involuntary microtremor even though not all of the literature agreed with the existence of the phenomenon or a correlation with stress. Using microprocessor technology, which offers the only hope of real-time analysis capability, a program was proposed to perform autocorrelations on taped voices from operational settings estimated to be stressful for the operator. The hypothesis was that, if measurable, the microtremor would vary in the degree of presence in some consistent relationship to the estimated level of stress prevailing upon the operator.

The decision to go directly to real life voiced output tapes for analysis was based on accepting a hypothesis that as stress increases, the vocal musculature microtremor, which is superimposed on F_0 as FM, changes with stress in some discernible relationship that can be detected and analyzed with instrumentation. There are ethical problems with creating high levels of stress in the laboratory and there are questions about the validity of simulated emotions, even as performed by skilled actors. Taped voice outputs from aircrews in life-threatening, high workload conditions were thought to have potential as good source materials for these studies. To this end, the USAF provided tapes of aircrews engaged in SEA air combat. Other live voiced outputs of pilots having inflight difficulties were obtained from the San Antonio Air Traffic Control facility. The lack of controlled variables and the individuality of each case reduces the methodological strength of such studies; however, the undisputed realism of the operational setting probably justified the use of such voiced outputs.

METHODS

The original plan to use government-supplied recorded voiced outputs from aircrews in combat was altered because the government-supplied recordings proved to be unsatisfactory for analysis. The tapes finally selected for test were of general aviation pilots having weather problems and receiving guidance from the San Antonio Air Traffic Control Tower. The difficulties with the combat tapes were mainly ones of excessive background noise and frequently unintelligible speech. Also, voiced outputs during combat engagements are usually very cryptic and rarely are of sufficient continuous speech duration for analysis, necessitating considerable splicing to eliminate pauses and to connect speech segments. Five seconds of continuous speech was selected as an appropriate tradeoff between the size of the statistical sample and time resolution in tracking stress change, i.e., the FM analysis was performed every one-half second, which gave 9 or 10 analyses on which to do statistics (means and standard deviations of each FM component).

The computer-based FM extraction system implemented for this study relied on waveform digitization as the data reduction technique. Data digitialization enhances the tests for intra- and inter-individual reliability, and for comparison with other stress data. In the waveform digitization approach, the FFT spectral analysis algorithm underlying the computational process was applied to the digitized representation of the waveform. The procedure is based upon an autocorrelation technique.

The front end portion of the system processed the speech by continually extracting the FM of the fundamental (F_0). The statistics portion computed a mean, square root, and standard deviation of the FM over the duration of the speech sample within the parameters of the 5 to 15 Hz frequency spectrum. The software also provided for printout of the statistical procedures carried out.

The Equipment

The equipment used in the voice stress project consists of (Fig. 1):

(1) a DEC PDP 11/34 computer running RT-11 software (GFE);
(2) a Computer Design and Applications MSP-3X array processor installed within the 11/34;
(3) a Krohn-Hite multimode filter model 3750;
(4) a Sharp RD-667 cassette deck; and
(5) a Wollensak, model 2820 AV, heavy duty cassette tape recorder.

The Sharp was selected as being adequate to preserve the original quality of the flight recordings. The Wollensak was to produce a third generation tape of 42 speech samples.

The Krohn-Hite is used as an adjustable filter and primarily used to prevent aliasing, which occurs when the original signal contains frequencies above one-half the sample rate (see any text on applying the discrete Fourier transform). The filter allows selectable rolloff rates (6, 12, 18 and 24 dB per octave) and for maximum anti-aliasing the highest roll rate was selected. The normal setup is 60 to 3000 Hz passband with maximum rolloff outside this range (24 dB per octave).

The PDP 11/34 contains a 12-bit (4096 counts) analog-to-digital converter. An existing and well tested software program developed previously by Technology Incorporated was adapted to operate the converter

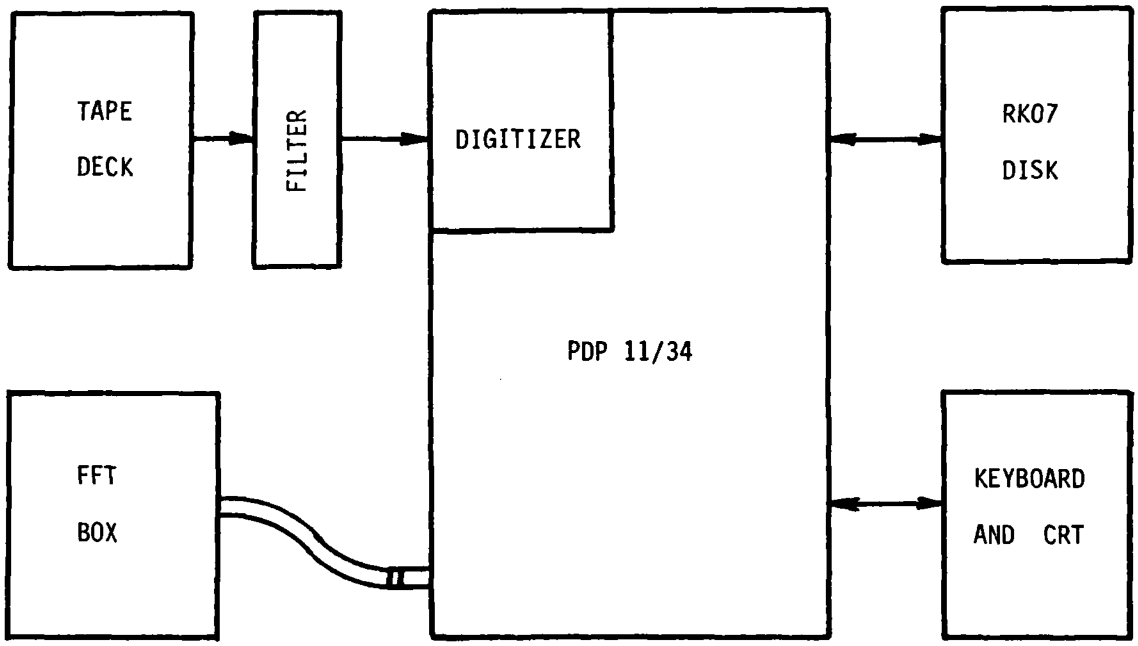


FIGURE 1. Voice Stress Assessment System for Processing and Analyzing Voice Output Data

in this application. The sampling rate can be varied from 16 Hz to 26,000 Hz. A 5000 Hz sampling rate was chosen largely by trial and error. The compromise was between waveform (speech) fidelity and disk space to hold the digitized speech. Another consideration was that for male speakers, there should not be appreciable harmonics above 2500 Hz. A higher sampling rate, therefore, reduces aliasing problems and improves waveform fidelity, and, further, increases disk space on which the digitized data are saved prior to further processing.

The MSP-3X is a low-cost array processor suited to performing Fast Fourier Transforms (FFTs). It was installed within the PDP 11/34 along with associated software provided by the manufacturer. This software allows the MSP-3X to be utilized from a FORTRAN program. The MSP-3X is capable of performing a 256-point autocorrelation in under 10 msec. This was felt to be fast enough to do voice stress analysis in real time, once the analysis algorithm was fully specified.

The autocorrelation is done via:

(1) a forward FFT of zero-padded raw sampled data (zero padding is the usual approach to doing an autocorrelation via the FFT, i.e., 128 samples of speech are padded with 128 zeros and a 256-point FFT taken);
(2) taking the power (sum of the squares of the reals and imaginaries); and
(3) an inverse FFT of the power yielding a sampled autocorrelation waveform.

The Software

The software developed by Technology Incorporated consists of an interactive package with three parts (data collection, real time analysis, offline analysis), and a program to generate test or synthesic frequency modulated (FM) waveforms (Appendix I). Programming of the real time analysis component was not accomplished. The scenario originally envisioned was to analyze digitized data offline until a technique was found which would measure stress. With a validated technique in hand, the array processor could then be programmed to analyze speech in real time (i.e., as the speech was being digitized). However, offline analysis as of this point in time has not yielded a valid technique to implement.

The system, as currently designed and operated has the capability to analyze data in real time, assuming a valid technique can be developed. From the standpoint of timing, an autocorrelation rate of 128 per second gives 8 msec per autocorrelation, which is adequate for real time:

Load array processor with 128 samples	.46 msec
Zero padded FFT	1.40 msec
Take power	1.00 msec
Inverse FFT	1.48 msec
Find peak of autocorrelation	.20 msec
	4.54 msec

It is assumed that the above steps would represent the major computation to be accomplished for a validated technique.

Further, the tradeoff is between increasing the number of cycles in order to get a better autocorrelation and being able to track changes in the fundamental. Consider: 50 msec is 1/20th of a second. If the fundamental is 200 Hz, this gives 10 cycles of the fundamental for analysis, which is enough for a reasonably good autocorrelation. However, if the fundamental has 20 Hz FM, then 1/20th of a second of data will have only one complete cycle of FM, which the autocorrelation will not be able to measure.

The package consists of a driver, prompting sections, data collection, and data analysis sections all written in FORTRAN. Subroutine libraries utilized were the FORTRAN library, the System library, a previously developed machine language data collection routine, and the array processor library. To use the package, one starts the driver which prompts for which of the three sections to use:

(1) The data collection section prompts for sampling rate, sampling duration in seconds, buffer size, channel number, and resulting disk file name.
(2) After validating the replies, it pauses to allow the operator to prepare the equipment and start the tape deck.
(3) Upon giving the computer a start signal, the voice signal is digitized. Afterwards, control returns to the driver allowing the user to select the next operation.

The analysis section prompts for about 20 parameters governing data analysis. Some of the major prompts are:

(1) Starting the stopping points (in seconds) within a data file.
(2) Autocorrelation rate (usually 100 or 128 per second of voice tape data).
(3) Autocorrelation size (this corresponds to the number of msec of data covered by the autocorrelation), "256" corresponding to 50 msec of data is the usual reply.
(4) The range of frequencies within which to search for the autocorrelation peak (typically 30 to 250 Hz for male speech).

The remaining prompts govern the nature and format of the results:

(5) The rate at which the FM analysis is done (usually twice per second).
(6) The range of FM frequencies to be covered (usually 5-15 Hz).
(7) Whether the autocorrelation periods are to be printed or displayed.
(8) Whether the FM results are to be printed or displayed.
(9) Whether the first autocorrelation waveform is to be printed or displayed.

After prompting is finished and validated, the computer outputs the complete set of parameters (Tables 1-2, Appendix II). See text in section 2.3 for further explanation.

The data analysis software was checked by performing the analysis upon data representing a sine wave with a controlled amount of FM. The analysis works as expected upon such data.

The Analysis Technique

Autocorrelations are computed over a span of collected data in an equally distributed fashion to yield a specific autocorrelation rate. For a given autocorrelation the appropriate section of collected digitized voice data is loaded into the array processor. A zero padded real forward FFT is performed. The results are multiplied times the complex conjugate of itself yielding the "power" or sum of squares of the reals and imaginaries. This occurs in the frequency domain. The result is inverse FFT transformed back into the time domain yielding a sampled autocorrelation waveform. (For those not familiar with frequency domain analysis, suggest: Applications of Digital Signal Processing, 1978, Prentice-Hall, Alan V. Oppenheim, Ed.).

The autocorrelation waveform is searched over a range corresponding to a fundamental frequency range for a peak. The location of the peak, the value of the peak, and its neighbors are unloaded from the array processor so that an interpolated peak may be found. The location of the peak corresponds to the period of the fundamental of the voiced speech. An interpolation is done to improve the accuracy of the period. A 3-point or parabolic interpolation is done. An interpolated peak occurring outside the search limits is bounded, the final interpolated period being that of the appropriate limit.

The autocorrelation peak was assumed to be a Gaussian or bell-shaped peak. From three samples near the peak the exact location of the peak can be determined. Since the top of the peak approximates a parabola, the corresponding peak calculation assuming a parabola can be done if the three points encompass the peak (one point to one side and two points to the other side of the exact peak) and if the peak is quite broad. The original intent was to implement parabolic interpolation first and then implement Gaussian interpolation if the peaks were found to be quite narrow. The peaks were found to be quite broad. The only difference between fitting a parabola and a Gaussian is taking logarithms of the three sampled amplitudes. It was originally thought necessary to make the pitch determination as accurately as possible and interpolation was the way to do this.

The interpolated periods are input for the FM analysis, i.e., if there is FM in the original voice signal, it will result in an up-and-down motion of the periods. The frequency of these up-and-down's will be the frequency of the FM. Thus, a Fourier analysis of the periods will extract the FM components. The periods are loaded into the array processor a section at a time, corresponding to the FM analysis rate and size. An ordinary forward real FFT is done and the power computed over the FM frequency range of interest. These powers are unloaded from the array processor to the PDP 11/34. They are normalized by their sum so that the fraction of power in a given frequency with respect to the total power over the frequency band of interest is printed. A moment is computed to give an indication of the relative location of the majority of the power. The mean and standard deviation of each frequency column is also computed for a given analysis run.

The analysis runs at about 1/4 real time with no effort to optimize the performance of the analysis software. Data collection runs in real time. There are no means to play back collected data (i.e., resynthesize the voice sound track from the disk files). No special treatment of noise, blank tape, or unvoiced speech was done. The averaging properties of the autocorrelation technique and the FM analysis are relied upon to reduce the effect of these factors. However, long blank sections of the tape were eliminated by dubbing them out.

Procedures

Twenty taped voice outputs of aircraft operators under various levels of inherent stress were obtained from civilian and military sources. The tapes were indexed and assessed as to aircraft operator stress levels (determined through subjective analysis) and audio quality.

Recordings obtained from the Air Traffic Control facility at the San Antonio International Airport were selected for analysis based upon their audio clarity, amount of displayed stress and the number of continuous speech samples. Three recordings were analyzed: two separate instances of aircraft operators lost in weather and one Air Traffic Controller assisting one of the pilots. In addition, two male voices in a non-stress environment were recorded and analyzed.

A Wollensak, model 2820 AV, heavy duty cassette tape recorder and a Sharp Educator, model RD-665 AV, cassette recorder were used in producing a third generation tape of 42 speech samples. The speech segments were sequentially ordered beginning with the operator's request for assistance to the tape's conclusion, the pilot's affirmed safety. Tapes were dubbed so that only one voice was recorded per tape. Silent spaces, pauses, etc., were dubbed out so that no more than one second of silence separated speech segments.

RESULTS

The data revealed the presence of FM (microtremor not confirmed) and its shift over the fundamental frequency (F_0). The failure to confirm microtremor was due to the presence of unavoidable noise on the tapes. The nature of digital signal processing is that for noisy signals, the results show a broad peak where in the ideal case, there would be a single value. In the case of the microtremor, the processing occurs in several steps.

The first step is the sampling and conversion of the speech signal. The Nyquist criteria requires that the sampling rate be at least twice the highest frequency of interest (in this case, the highest order harmonic of the speech pitch).

Having sampled and digitized the signal at a sufficient rate to encompass the desired harmonics, pitch determination is done. The effect of noise in the original signal is to introduce greater variability in the computed pitch (i.e., a less accurate pitch). The software actually computes the period which is the reciprocal of pitch. Parabolic interpolation was an effort to improve the speech period determination. In the presence of noise, the parabola is flatter and interpolation is less accurate. At a sampling rate of 5 kHz, uninterpolated period accuracy would be 200 msec. The signal-to-noise ratios in the voice tapes are such that even this accuracy is unobtainable. Assuming a 200 Hz pitch, or equivalently a 5000 μsec period, would, without noise, yield a 1:25 accuracy. For those parts of the voice tape that gave pitch results, a 1:10 accuracy appears more reasonable.

The last stage of the analysis is the FM determination of speech periods. Assuming a 150 Hz fundamental with 10% accuracy, we get a 15 Hz FM band due to noise. This noise in the FM band of interest, which is added to the FM due to microtremor, tends to disguise the microtremor under these conditions. The hypothesis that the microtremor present in the frequency range of 5-12 Hz will vary consistently with changing levels of stress, could not be tested, since the microtremor could not be confirmed using this technique. The results are confirmed to reporting the peak pitch periods that were brought out in the autocorrelation.

Tables 1 and 2 portrayed the format for data analysis display. The analysis program delivered two printouts for each analysis segment. The first listing is that of all the autocorrelations for each 30-sec segment of analyzed vocalizations. The second printout, the FM analysis, is a listing of the peak period means for each frequency bin between 5-11 Hz. The bin containing the most FM (highest mean) is considered the frequency of interest for that particular speech segment.

The pilots and controller vocalizations are printed in tables in order of analysis so that speech segments and times of the peak mean can be seen together. Each table is explained in the text that follows.

For the purpose of making comparisons between vocalizations obtained under real life, stressed conditions and those obtained in unstressed familiar surroundings, vocalizations were obtained from two adult male subjects during a relaxed recording session. Subject No. 1 was an audiovisual professional with speech training, while Subject No. 2 had no formal speech training. A summary of the analysis of two 30-sec speech segments for each subject is presented in Table 3. As can be seen in this table, both speakers have their FM peak means in the 5 Hz frequency bin, which differs considerably from the airborne and controller responses.

The peak means for the distributed FM for the pilot of Cessna 30-Golf are shown in Table 4. The highest peak means occur during initial contact with the ground control (Thrush Control), a period when anxiety would be expected to be high.

By analysis, the initial peak mean is in the 7 Hz frequency bin for the first 30 sec of vocalization and in the 6 Hz frequency bin for the second 30-sec period of analysis. For the next six analysis segments, the peak means all fall in the 5 Hz frequency bin. The peak means rise again to 7 Hz in analysis segments nine and ten (240-299.06 sec) where the pilot is speaking with considerable emphasis, however, the effect of a background voice on the analysis segments is unknown.

The peak means from the distributed FM for the pilot of Cessna 48-Golf are shown in Table 5. Unlike the 30-Golf tape, the peak FM means remain at or above the 7 Hz level throughout the taped vocalizations. Only in the eighth analysis segment (210-239.06 sec) does the peak mean fall to 6 Hz; however, in the final analysis segment, it falls to a 5 Hz frequency bin, which may or may not be a reliable value due to interference. The frequency bins with very low values (i.e., 430-439.38 sec) can be discounted, probably due to vocalizations too short in duration, or no vocalization at all.

When comparing 30-Golf and 48-Gold peak means, 48-Golf has peak means that are consistently in higher frequency bins than 30-Golf, however, in the absence of baseline recordings for the two pilots, nothing definitive can be said about these differences. On the basis of subjective impression of the vocalizations, if a higher peak mean is indicative of higher emotional levels, then 30-Golf should have registered peak means in higher frequency bins than the analysis showed, since the vocalizations sounded more stressed than the numerical values would seem to indicate, assuming, that is, that peak means of around 5 Hz represent relatively unstressed vocalizations.

The peak means for the distributed FM for the Air Traffic Controller working 30-Golf shows more variability than either of the two pilot tapes (Table 6). For the first 3-1/2 minutes of tape analysis, the mean peak FM is located in frequency bin 5 Hz, but then it rises to the 6 Hz bin for one analysis segment before falling back to the 5 Hz bin. This fluctuating pattern continues until the 18th analysis segment (930-959.06 sec) where the peak FM rises to the 7 Hz bin, falls back to 6 Hz for one segment, then spikes to the 9 Hz level before finally settling back to the 5 Hz frequency bin for the rest of the recording. There is no accounting for this range of variability, except on a subjective evaluation of the emotional state of the controller based on his vocal output.

The fluctuation of FM in the 5-9 Hz range certainly indicates that a measurable change is occurring that might relate to some stress effects. Again, in the absence of baseline data on the individuals, it would be very difficult to quantify. Future studies could examine voices from operational settings, but it would be desirable to have previous laboratory data on the same individuals.

The same mean peak FM data were graphed as FM in Hz over time and as FM normalized (percentage of one) over time. The data for the pilot of 30-Golf (Fig. 2) clearly show the same relationship as was depicted in Table 4 for the FM charted over time. The normalized data for 30-Golf produces a far more graphic picture of the variability of the FM over time (Fig. 3).

For 48-Golf, the same relationship holds in the mean peak FM over time (Fig. 4) and in the normalized data (Fig. 5). The variability, both as a percentage and as peak FM frequency change is even more evident than in the other pilot. Unfortunately, no baseline data exist for either pilot, so the change from normal, unstressed vocalizations cannot be determined.

The Air Traffic Controller vocalization data make a very clear picture of the variability of the peaked means of the FM (Figs. 6 and 7). Despite the obvious and dramatic variability in the Controller's FM distribution, its correlation with a level of stress or emotional status is purely subjective, as it was for the pilot data.

DISCUSSION

The vocalizations analyzed in this effort were collected from real-world stressor situations. Although the data are compared between pilots, controllers and unstressed male speakers, it is felt that the useful value should have been a comparison between the speakers and their own baseline.

The failure to find microtremor was due to noise levels in the tapes that could not be economically filtered or averaged out using the capability and techniques reported here. This made testing the hypothesis impossible; however, it was possible to continue to extract pitch information. This provided some information about differences between speakers and further confirmed the hardware and software capability.

The data collected here does tend to support findings that as emotional stress mounts, the speech signal has a tendency to distortion and displacement into the higher frequencies. Popov (1971) found this to be true as he reports on his Russian work. There is, unquestionably, a difference between speakers. The two unstressed, male speakers showed no tendency toward peak mean FM distributions as was seen in the pilot and controller data. Between pilot differences are apparent, as are differences between either pilot and the controller, at least as far as the degree of variability in FM peak mean distributions.

The significance of the variability and its degree is less clear. Given the extreme range of individual responses to stress, it does not seem possible to make judgments of human response potential based on one-time analyses of vocalizations without a reliable baseline, whether that baseline is for the subject under analysis, or is a human performance data base. Certainly, for the combat pilot, the need for establishing

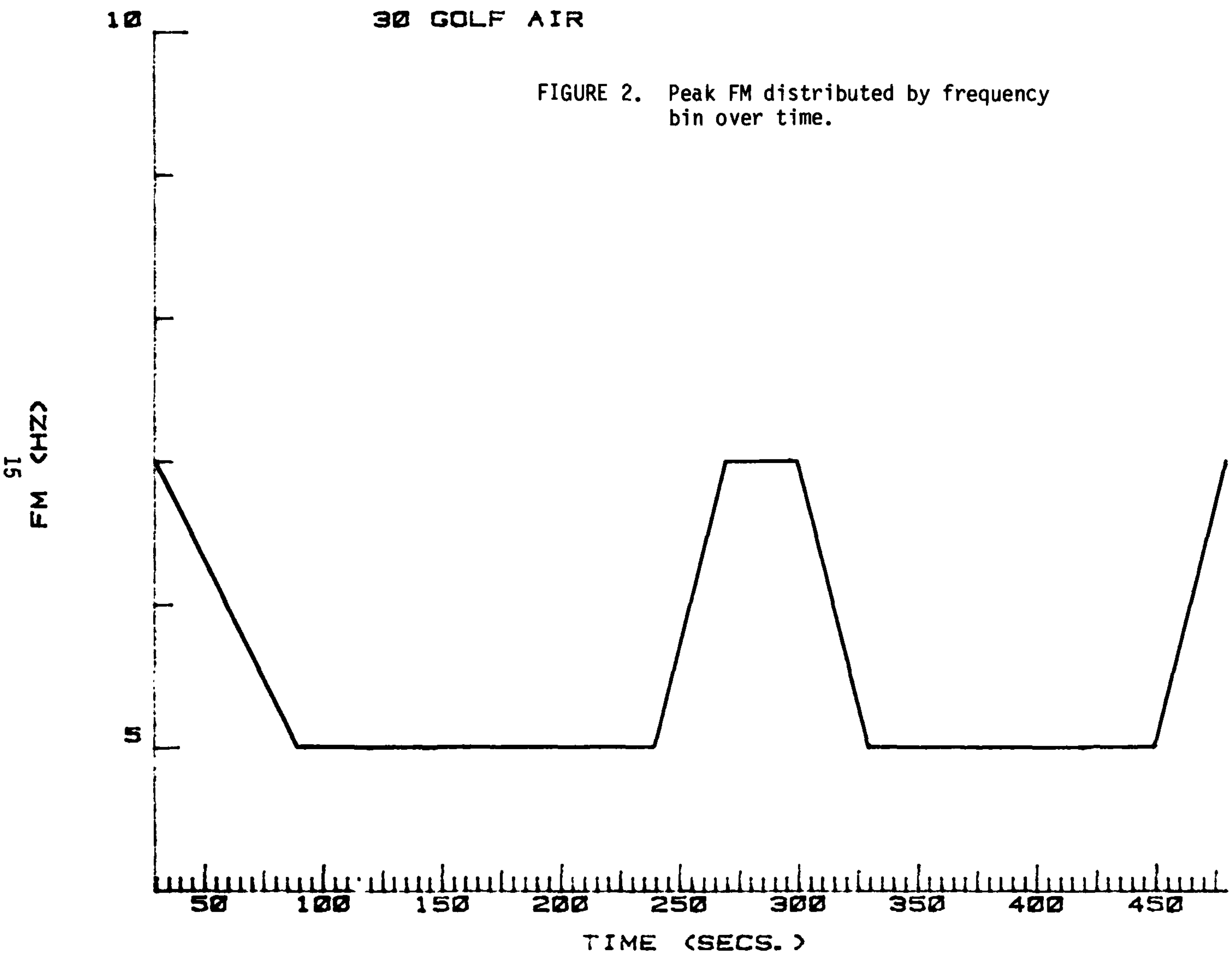


FIGURE 2. Peak FM distributed by frequency bin over time.

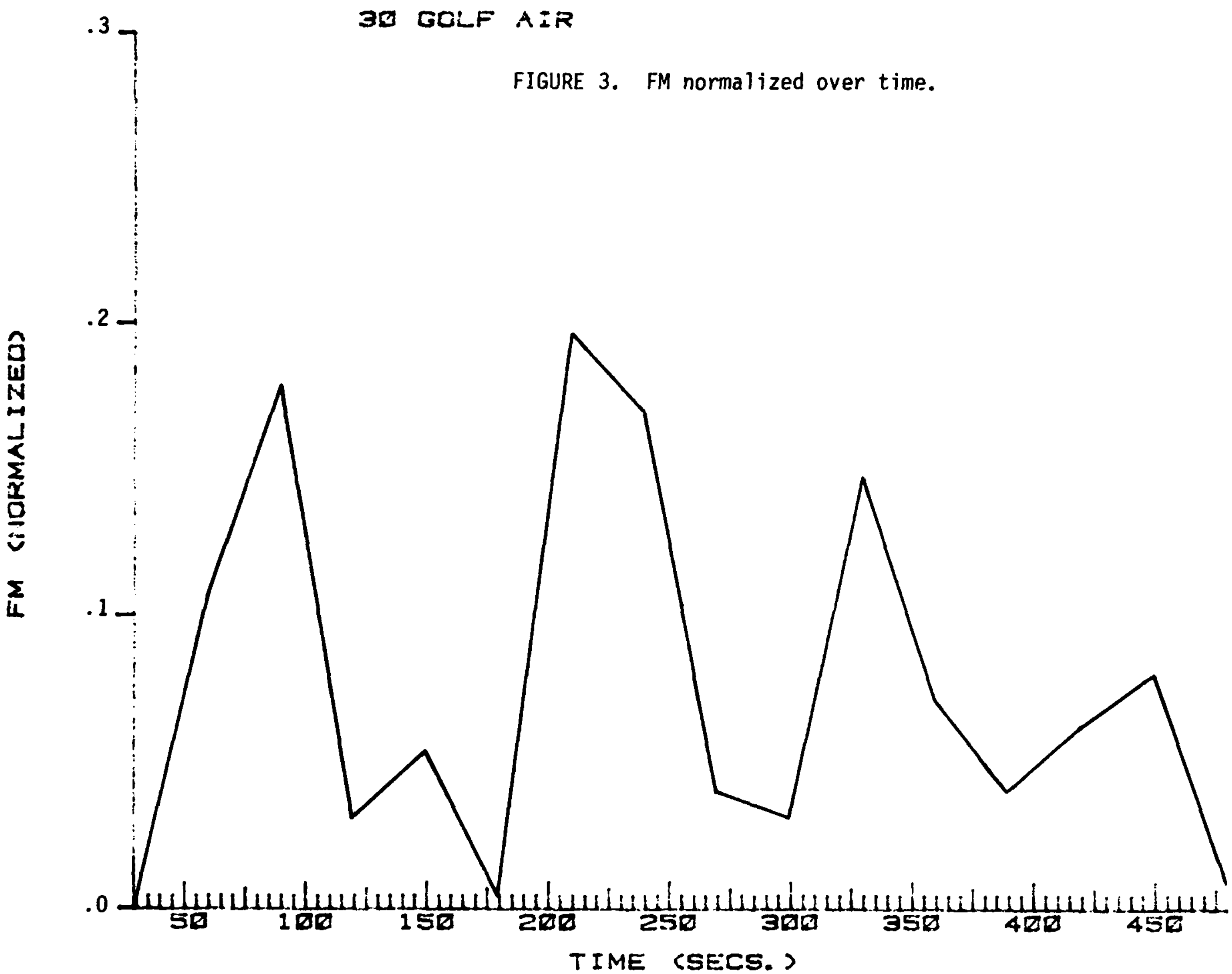


FIGURE 3. FM normalized over time.

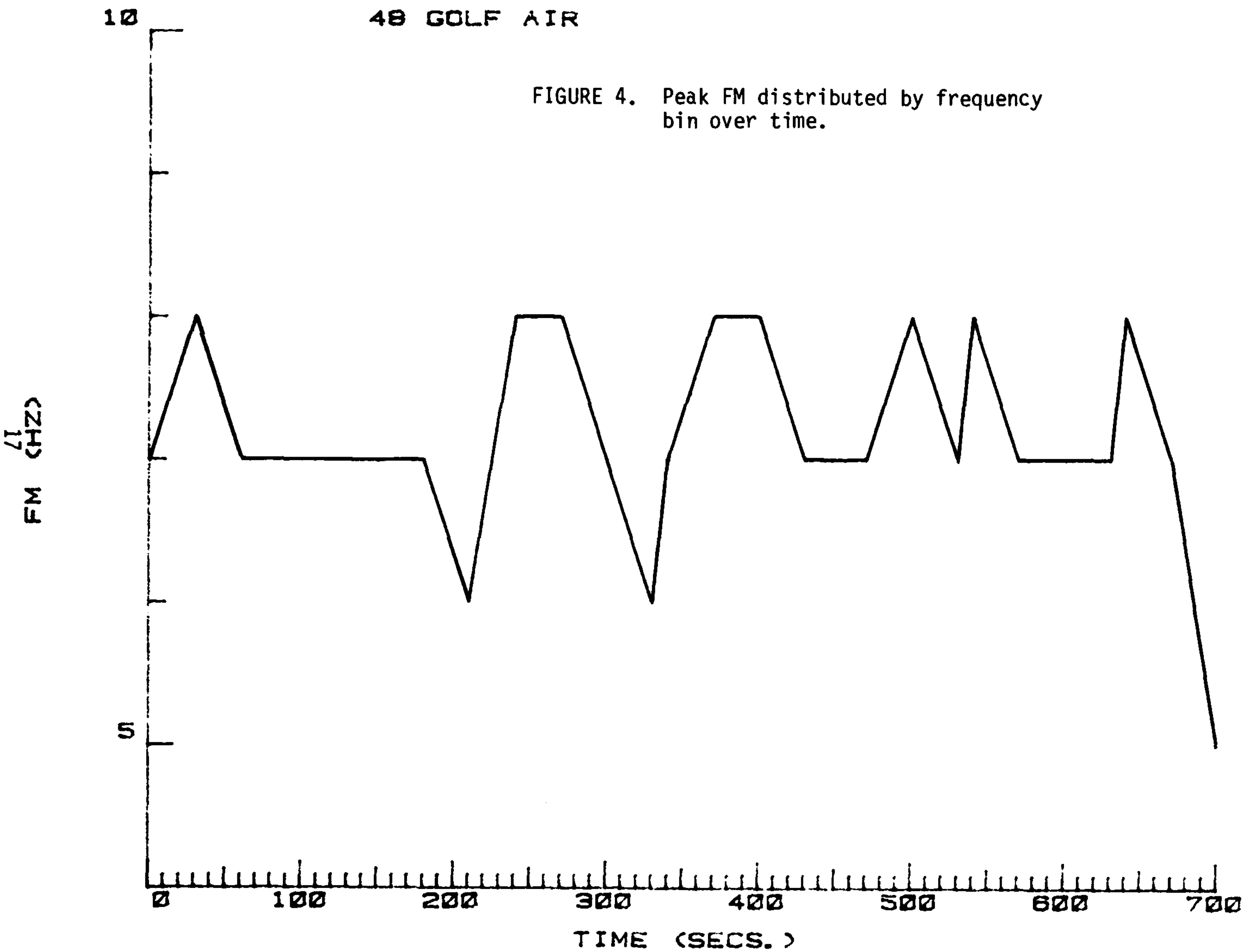
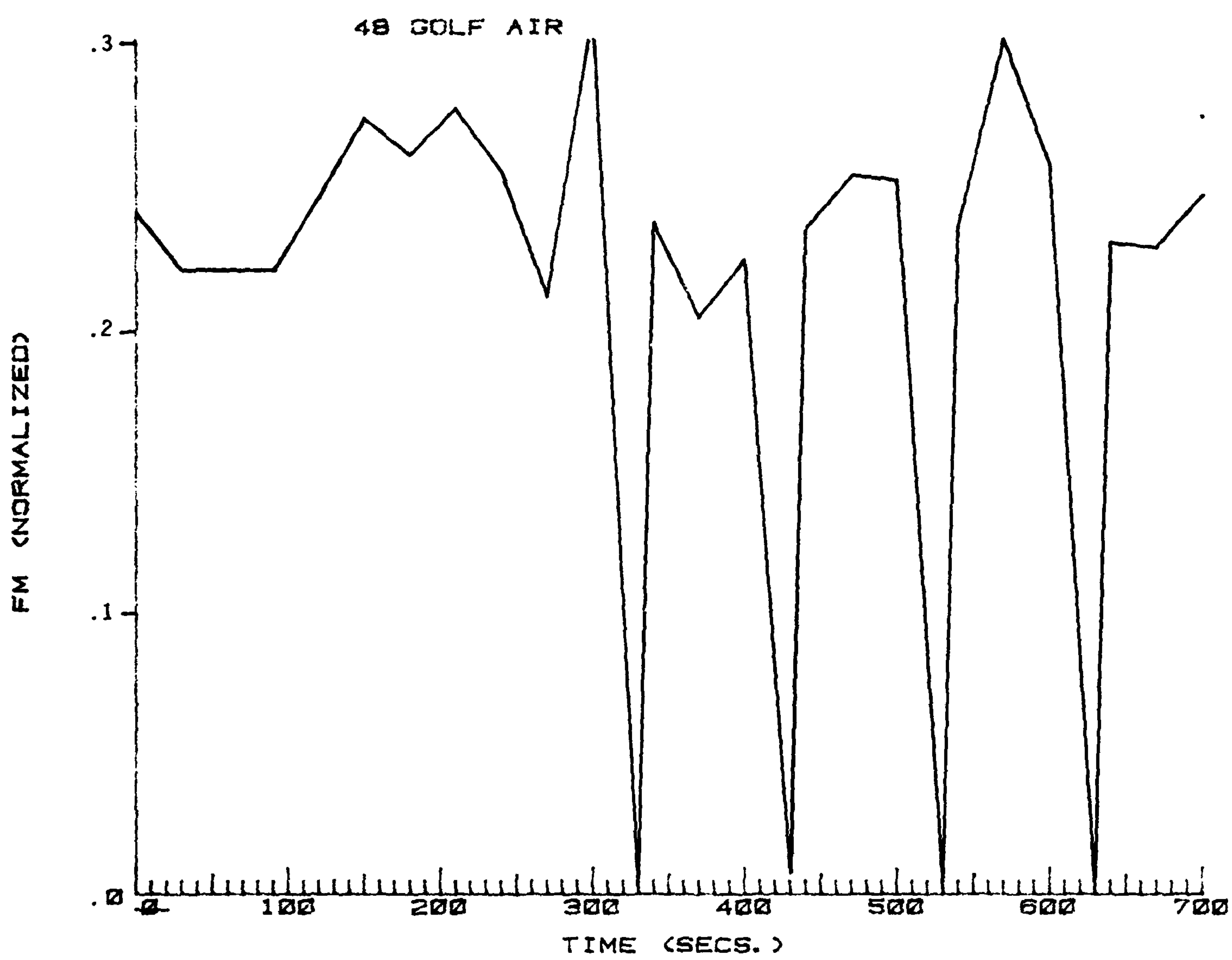


FIGURE 4. Peak FM distributed by frequency bin over time.

FIGURE 5. FM normalized over time.



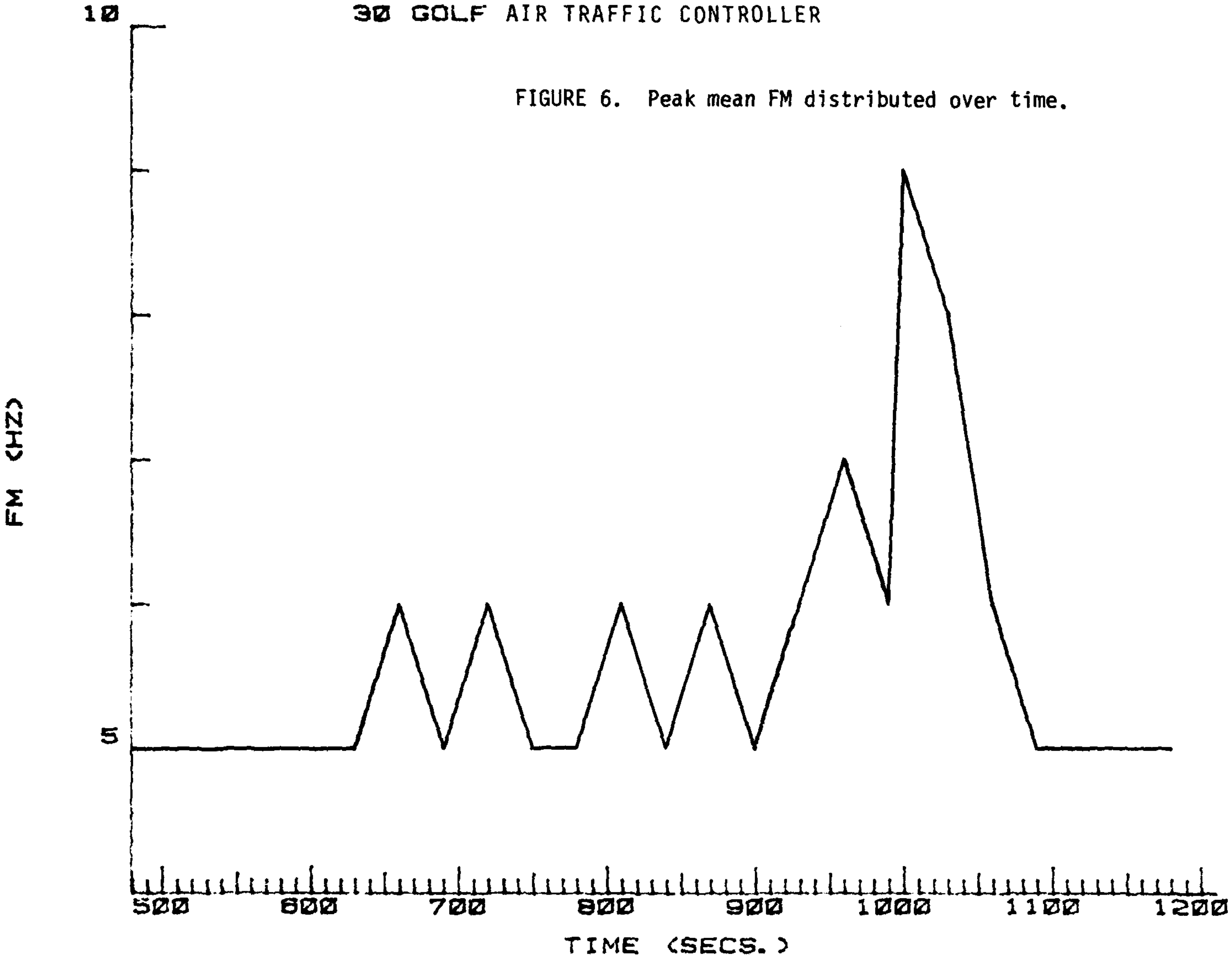
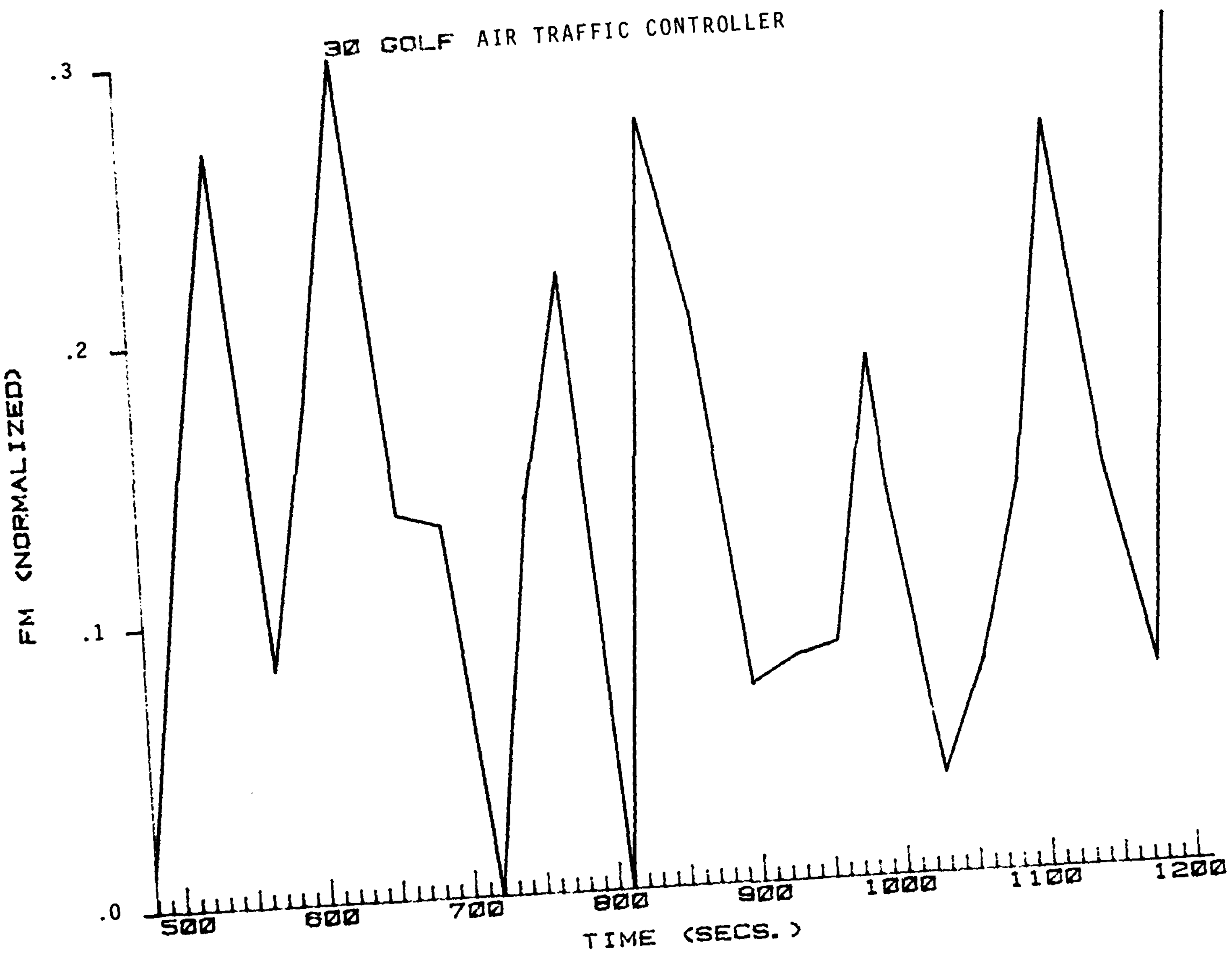


FIGURE 6. Peak mean FM distributed over time.

FIGURE 7. FM variability normalized over time.



a new baseline would seem to be a daily need. The thesis that any stress probably effects some response, whether that change is currently measurable or not, is probably a valid observation. The thesis that stress affects performance in a detrimental way is not always true. To a degree, stress can actually improve task performance, in some cases. An organism's response to stress varies considerably from time to time and over a broad range of effect. This variance can be considerable, yet remain below a level of measurable performance interference. In fact, in some instances, performance may, for all appearances, remain normal until the moment of catastrophic collapse. On an individual basis, an organism has stress tolerance limits that vary according to the daily state of the organism. A pilot, who performs superbly on a given mission, may make frequent mistakes on essentially the same mission when suffering the effects of a sleepless night due to gastric distress, or a reprimand from the boss. Emotional and physical pressures during combat are often severe and protracted. The emotional effects of one particularly hazardous mission may linger on for days, or the cumulative effects of 7-day-a-week operations may finally culminate in an accident that is of a totally unexpected variety, e.g., landing gear up. To be of any real value in a combat environment, it would be necessary to know at any given time where the pilot's stress state is in relation to his own baseline, assuming also, that the baseline tells something about the individual's stress tolerance. In particular, it would be most valuable to know his state prior to mission launch, because once airborne and inside enemy territory, recall on an individual basis is unusual, if not impossible. Severe stress in aerial combat usually lasts only a bare few minutes, sometimes seconds, and the opportunity to intervene on behalf of an overstressed individual is rare.

The data collected during this effort did not meet expectations for at least several reasons: lack of sufficient instrumentation and capability to control some variables.

The procedure used in the measurement of the speech segments needs to be improved. Additional programming is suggested to mark and store the start and stop parameters for each speech sample. The variation in the analysis and the inconsistent repeatability of the data, could be due to the lack of instrumentation to assure that the analysis of the voice signal begins with the onset of vocalization and concludes precisely at the end of the speech segment. It is recommended that the voice output being digitized on a disk also allows the computer operator to hear the speech segments as they are processed by the computer. To augment this procedure, an oscilloscope could provide additional feedback in securing accurate start and stop times for the vocalizations.

There should be careful preparation of the voice output tapes used for analysis. Extensive conditioning (clean-up) was not done on the tapes analyzed in this study. A conditioning procedure should be used on all tapes to be analyzed, even if to the ear, clarity of the tape seems apparent. A 60 Hz hum, frequently found as background noise imbedded in the speech sample, can filter out the fundamental (F_0). Although caution must be taken to avoid erasing certain overtone series of the speech frequencies, certain hums and ambient noise can be removed from a tape without endangering the speech spectrum. It is probable that more filtering can be done with the software program.

The equipment used in recording the tape for analysis was not of the same instrumentation or quality for all tapes. Uniformity of equipment should be used in duplicating speech segments. Tape recorders should be of industrial quality. They should be tested as to degree of wow and flutter as flutter can be greater than FM, thereby masking it out. These were dollar resource problems.

The PDP 11/34 should be equipped with audio playback and also be capable of graphic display of data. The 12-bit A-to-D converter should be replaced with a 16-bit A-to-D converter; apparently, the 12-bit is not quite capable enough for the job.

Research results supported the program's instrumentation ability to identify the divergency of the FM from the F_0 and specify its frequency component between 5-11 Hz; however, there is insufficient information present to firmly relate the FM activity to stress. For the present, at least, it does not appear that we have the technology or insight to do meaningful real-time stress analysis of nonlaboratory voiced outputs.

In summary, the hypothesis that a microtremor was present in the voice which varied in relation to stress, could not be tested because the microtremor could not be found using the autocorrelation technique. The failure is thought to be due to noise in the 5-11 Hz FM frequency band of interest. Pitch was extracted and peak period means recorded, which tended to shift primarily between 5-9 Hz. This was not relatable to any subjectively determined level of stress in the speaker's voice. In our opinions, subjective determination of stress will be near impossible without the evaluator being familiar with the speaker. To judge voiced outputs recorded from an operational setting will always be particularly difficult because of the context of aircraft operations, i.e., is the operator yelling to overcome interference, gain attention, because he is excited or because he is scared? So, given the difficulty of identifying vocal correlates of emotional states in operational environments, it is recommended that future research concentrate on developing laboratory baselines and then collecting operational data, with the operator reporting his stress level. Autocorrelation technology may not be the vehicle for analysis. [Conversations with an industry leader in voice analysis (Signal Technology, Inc., 1982) reports that statistical summaries of period and amplitude are the best they can do on continuous vowels recorded under laboratory conditions. Their autocorrelation technique for extracting pitch is the same as used on this effort.]

RECOMMENDATIONS

A possible direction for further research might be to work toward the development of a comprehensive FM stress scale. Research areas to be included might be the study of (1) FM characteristics of amplitude and periodicity, and (2) the FM relationship to the following vocal parameters of the speech segment: (a) fundamental frequency (F_0)--its median and range and, most importantly, its contour vs. time; and (b) the energy distribution in the spectrum, particularly between 500 and 1000 Hz. At this time, research has indicated that these acoustical cues are of primary importance in the communication of essential information regarding emotional expression.

REFERENCES

1. Williams, C.E. and K.N. Stevens. 1972. Emotions and speech: some acoustical correlates. J. Acoust. Soc. Amer. 52:1238-1250.

2. Hecker, M.H.L. 1971. Speaker recognition: an interpretive survey of the literature. ASHA Monographs No. 16, American Speech and Hearing Assoc., Washington, D.C.

3. Lynch, G.E. 1934. A phonophotographic study of trained and untrained voices reading factual and dramatic material. Arch. Speech 1:9-25.

4. Fairbanks, G. and W. Provonost. 1939. An experimental study of the pitch characteristics of the voice during the expression of emotion. Speech Monogr. 6:87-104.

5. Fairbanks, G. and L.W. Hoaglin. 1941. An experimental study of the durational characteristics of the voice during the expression of emotion. Speech Monogr. 8:85-91.

6. Lieberman, P. 1961. Perturbations in vocal pitch. J. Acoust. Soc. Amer. 33:597-603.

7. Lieberman, P. and S.B. Michaels. 1962. Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech. J. Acoust. Soc. Amer. 34:922-927.

8. Fonagy, I. and K. Magdies. 1963. Emotional patterns in intonation and music. Z. Phonetik, Sprachwissenschaft und Kommunikationsforschung 16:293-326.

9. Uldall, E. 1960. Additudinal meanings conveyed by intonational contours. Lang. Speech 3:223-234.

10. Skinner, E.R. 1935. A calibrated recording and analysis of the pitch, force and quality of vocal tones expressing happiness and sadness; and a determination of the pitch and force of the subjective concepts of ordinary, soft and loud tones. Speech Monogr. 2:81-137.

11. Huttar, G.L. 1968. Relations between prosodic variables and emotions in normal American English utterances. J. Speech Hearing Res. 11: 481-487.

12. Hollien, H., J. Michel and E.T. Doherty. 1973. A method for analyzing vocal jitter in sustained phonation. J. Phonetics 1:85-91.

13. Beckett. R.L. 1969. Pitch perturbation as a function of subjective vocal constriction. Folia Phoniat. 21:416-425.

14. Rozsypal. A.J. and B.F. Millar. 1979. Perception of jitter and shimmer in synthetic vowels. J. Phonetics 7:343-355.

15. Lippold, O. 1971. Physiological tremor. Sc. Am. 224:3.

16. Edson, R.K. 1976. The Dektor psychological stress evaluator as a research instrument. Master's Thesis submitted to the National Graduate University.

17. Bell, A., W.H. Ford and C.R. McQuiston. 1972. Product information: psychological stress evaluator. Psychiat. and Neurol.

18. Dektor. 1971. The psychological stress evaluator. Dektor C.I.S. Inc., 5508 Port Royal Rd., Springfield, VA 22151, USA.

19. McGlone, R.E., C.R. Petrie and Frye. 1974. Acoustic analysis of low-risk lies. J. Acoust. Soc. Amer. 55:S20. (Abstract).

20. Papcun, G. 1974. The effects of psychological stress on speech: literature survey and background. J. Acoust. Soc. Amer. 55:422-423 (Abstract).

21. Lambert, R.W. 1974. The psychological stress evaluator: a recent development in lie detector technology. In Problems in Law and Medicine, University of California, Davis, Law Review, pp. 332-350.

22. Smith, G.A. 1977. Voice analysis for the measurement of anxiety. Brit. J. Med. Psychol. 50(4):367-373.

23. Eden, G. and G.F. Inbar. 1975. Interim Report Res. No. 050-334, Technion, Haifa, Israel.

24. Eden, G. and G.F. Inbar. 1976. Psychological stress evaluators: EMG correlation with voice tremor. Biol. Cyber. 24(3):165-167.

25. Eden, G. and G.F. Inbar. 1978. Physiological model analysis of involuntary human-voice tremor. Biol. Cyber. 30(3):179-185.

26. Popov, V.A., P.V. Simonov, M.V. Frolov and L.S. Khachaturyants. 1971. Frequency spectrum of speech as an indicator of the degree and nature of emotional stress. Zhurnal Vysshey Nervnoy Deyatelnati 1:104-109. Reproduced by NTIS, PRS 52698, Springfield, VA.

27. Personal communication, 1982, Signal Technology, Inc.

APPENDIX I

SOFTWARE INDEX

A)	SINGEN.FOR	Program to generate disk file having same format as voice data with operator specified FM frequency and depth of modulation.
B)	VSA.FOR	Driver for voice stress software.
C)	VSACOM.FOR	Chain or common area for voice stress software. Holds data collection and analysis parameters.
D)	VSACDD.FOR	Driver and prompting routine for data collection.
	a) .MAIN.	Driver.
	b) MAIN11	Prompting routine.
E)	VSAATD.FOR	Driver and supervisor routine for data collection.
	a) .MAIN.	Driver and storage allocator.
	b) IATD	Buffer manager.
	c) RACMPT	Buffer overrun (completion routine).
F)	IATD.MAC	Assembly language A-to-D software.
	a) IAATD	Setup routine (FORTRAN callable).
	b) MCATD	Multichannel interrupt routine (not used).
	c) ATD	Single channel interrupt routine.
	d) PARMS	Parameter validation of FORTRAN call.
G)	VSARTA.FOR	Real time voice stress analysis (stub).
H)	VSAADD.FOR	Off-line voice stress analysis (data on disk).
	a) .MAIN.	Driver.
	b) ANAL	Data analysis using array processor.
I)	MYLIB.FOR	Prompting for VSAADD.FOR.
	a) PROMPT	Voice data analysis prompting.
	b) ISPWR2	Check to see if a power of 2.
	c) SCROLL	Turn scrolling on or off.
	d) CLR	Clear screen.
	e) CURSOR	Position the cursor.
	f) DELAY	Delay specified number of seconds.
	g) IIRF50	File name prompt in read mode using cursor positioning.
	h) IFPMT	Prompt for file name using cursor positioning.
	i) NYCHG	Allow selective re-execution of prompts with cursor positioning.
	j) IQ2	Prompt for integer using cursor positioning.
	k) IIQYN	Prompt for yes/no answer using cursor positioning.
	l) IIRQ	Prompt for real value using cursor positioning.
J)	PAGES.FOR	Output formatting and statistics.
	a) PAGE1	Autocorrelation periods output.
	b) PAGE2	FM content output, includes moment calculations and statistics.
	c) PAGE3	Output means and variances.
	d) GRAPH	Rough graph output of FM content.
K)	BORDER.FOR	Subroutines for CRT terminal fixed display.
	a) MARQUE	Put up fixed text.
	b) BORDER	Draw border around screen.
	c) LFT	Draw line from left to right.
	d) RT	Draw line from right to left.
	e) ASCEND	Draw line from bottom to top.
	f) DESCND	Draw line from top to bottom.
L)	PTLIB.FOR*	Previously developed prompting routines (IQ,IRT,IQYN, etc.).
M)	FORLIB.FOR*	FORTRAN library (SIN,COS,SQRT, etc.).
N)	SYSLIB.FOR*	RT-11 library (PRINT,IREADW, etc.).
O)	MSPLIB.FOR*	Array processor library (ZRFFT,RIFT,CCMXT, etc.).

* Source listings not available or not included.

APPENDIX II - FIGURES

TABLE 1. Sample listing of autocorrelation periods.

TABLE 2. Sample of analysis printout for peak FM by frequency bin. Peak FM for this 30-sec analysis is in the 5 Hz frequency bin.

TAPE: <u>BASELINE</u>

AUTOCORRELATION PERIOD MEANS
FREQ BINS 5-11 Hz

5 Hz	6 Hz	7 Hz	8 Hz	9 Hz	10 Hz	11 Hz	
<u>.219</u>	.178	.115	.090	.075	.085	.067	Baseline Subject #1
<u>.249</u>	.166	.117	.092	.084	.079	.051	
<u>.228</u>	.176	.242	.301	.206	.250	.213	Baseline Subject #2
<u>.179</u>	.167	.115	.099	.074	.077	.058	

TABLE 3. Summary of vocalizations for two baseline subjects shows mean peak FM occurring in the 5 Hz frequency bin.

TAPE: 30 Golf-Air

TABLE 4. Peak FM autocorrelation period means for pilot of Cessna 30-Golf. Tape contains only the pilot's voice.

SECONDS INTO TAPE	AUTOCORRELATION PERIODS MEANS FREQUENCY BINS 5-9 Hz: 5 Hz	6 Hz	7 Hz	8 Hz	9 Hz	VOCALIZATIONS
0.00-29.06	0.102	0.103	<u>0.113</u>	0.098	0.070	"Thrush Control, this is uh...Cessna Three-Zero Golf. I'm thirty miles east of town. I'm got caught up on top and I can't get down. I'm circlin' in a hole." "Ten thousand-five hundred feet." "Thrush Control, you still read me?" "Two-four ze-
30.00-59.06	0.097	<u>0.137</u>	0.096	0.106	0.111	ro ? uh...pardon me, I'm on a sixty degree radial out of San Antone VOR." "Thrush negative!" "I'm on a heading of uh...two-zero-two-zero circling." "This is Three-Zero Golf, you still read me?"
60.00-89.06	<u>0.153</u>	0.098	0.097	0.083	0.098	"I've...hear you fine now." "I'm at ten thousand feet on a zero-six degree radial out of San Antonio VOR, and I'm about thirty to forty miles out." (Voice of Ground Control in background) "Golf, my heading is...due north." (Voice of Ground Control in background)
90.00-119.06	<u>0.120</u>	0.110	0.095	0.119	0.103	"Roger, but I'll be in the clouds." "That's uh...negative." Three hours of

CONTINUED...

TAPE: 30 Golf-Air

Table 4a.

SECONDS INTO TAPE	AUTOCORRELATION PERIODS MEANS FREQUENCY BINS 5-9 Hz 5 Hz	6 Hz	7 Hz	8 Hz	9 Hz	VOCALIZATIONS
120.00-149.06	0.125	0.105	0.084	0.110	0.104	fuel on board and uh...I'm in a big hold, I can circle for awhile." "Thrush VOR, this is uh...Cessna Three-Zero Quebec. This turbulence is too hard up here, I gotta' come down."
150.00-179.06	0.114	0.104	0.089	0.084	0.100	"Golf!" "No instrument rating nor equipped." "Roger!" "No thunderstorms
180.00-209.06	0.157	0.100	0.066	0.089	0.090	in the way here." "San Antone VOR, uh...I am in the clouds." "Three-Zero Golf, two souls aboard."
210.00-239.06	0.151	0.111	0.105	0.089	0.073	"Cessna One-Seventy-Two." "Locate me!" " ? for five seconds holding down."
240.00-169.06	0.103	0.110	0.122	0.099	0.104	"Affirmative!" "Luling!" "Two-four-zero." "Thrush positive!"

CONTINUED...

TAPE: 30 Golf-Air

Table 4b.

SECONDS INTO TAPE	AUTOCORRELATION PERIODS MEANS FREQUENCY BINS 5-9 Hz 5 Hz	6 Hz	7 Hz	8 Hz	9 Hz	VOCALIZATIONS
270.00-299.06	0.119	0.100	<u>0.120</u>	0.119	0.100	"Thrush, I'm now on a zero-five-six-zero radial out of San Antone VOR." "<u>Give me La Vernia</u> VOR uh...VOR number, and I'll get <u>a call check</u> for you." (Voice of Ground Control in background)
300.00-329.06	<u>0.146</u>	0.097	0.103	0.104	0.094	"Roger!" "Got a two-ten radial out of La Vernia VOR." "Affirmative!"
330.00-359.06	<u>0.129</u>	0.119	0.111	0.113	0.114	"Two-zero degrees." "I'm on a heading of two-ten." "I'm coming down to San Antone VOR at two-three-five heading."
360.00-389.06	<u>0.122</u>	0.115	0.093	0.089	0.077	"Negative." "Roger!" "Fifteen minutes."
390.00-419.06	<u>0.127</u>	0.121	0.097	0.099	0.091	"Three-Zero Golf, my approach on VFR now." "Affirmative!" "Uh...just a minute, I'll try to find me."

CONTINUED...

TAPE: 30 Golf-Air

Table 4c.

SECONDS INTO TAPE	AUTOCORRELATION PERIODS MEANS FREQUENCY BINS 5-9 Hz					VOCALIZATIONS
	5 Hz	6 Hz	7 Hz	8 Hz	9 Hz	
420.00-449.06	<u>0.131</u>	0.105	0.100	0.098	0.074	"I'm just approaching...I don't know just uh...I'll find out in a minute and I'll tell you." "Roger!" "Control, this is uh...Cessna."
450.00-479.06	<u>0.115</u>	0.113	<u>0.115</u>	0.097	0.083	"Three-Zero Quebec. I'm north of uh...northeast of San Marcos approaching now the old airport up here."

TAPE: 48 Golf-Air

TABLE 5. Peak FM autocorrelation period means for pilot of Cessna 48-Golf. Tape contains only pilot's voice.

SECONDS INTO TAPE	AUTOCORRELATION PERIODS MEANS FREQUENCY BINS 5-9 Hz					VOCALIZATIONS
	5 Hz	6 Hz	7 Hz	8 Hz	9 Hz	
0.00-29.06	0.095	0.096	<u>0.133</u>	0.092	0.075	"San Antonio approach, this is Cessna Eight-One-Four-Eight Golf." "Zero-four-three-seven, altitude nine-five."
30.00-59.06	0.107	0.083	0.115	<u>0.122</u>	0.089	"That's correct." "Four-Eight Golf. Naw, that's negative." "Four-Eight Golf."
60.09-89.06	0.110	0.121	<u>0.122</u>	0.105	0.089	"Negative!" "Four-Eight Golf."
90.00-119.06	0.104	0.100	<u>0.122</u>	0.096	0.085	"We've got about two and half hours." "Four-Eight Golf from Denver." "Four-Eight Golf.
120.00-149.06	0.094	0.098	<u>0.136</u>	0.107	0.067	"Uh...is there any possibility of getting into San Antonio at all? I'm familiar with the airport and rather go in there than take a chance." "If I can get down good, you say you can get me in?"
150.00-179.06	0.115	0.085	<u>0.151</u>	0.118	0.077	"O.K. I'll fly over the city and see what happens." "Four-Eight Golf." "Four-Eight Golf."

CONTINUED...

TAPE: 48 Golf-Air

Table 5a.

SECONDS INTO TAPE	AUTOCORRELATION PERIODS MEANS FREQUENCY BINS 5-9 Hz: 5 Hz	6 Hz	7 Hz	8 Hz	9 Hz	VOCALIZATIONS
180.16-209.06	0.116	0.095	0.144	0.093	0.097	"I've got a...looks like a hole off to the left over here. I'm gonna' circle back to the left." "I'm at nine thousand." "Four-Eight Golf at seven thousand."
210.00-239.06	0.097	0.153	0.098	0.123	0.087	"Four-Eight Golf." "Four-Eight Golf."
240.00-269.06	0.125	0.083	0.110	0.141	0.131	"Four-Eight Golf. We're still barely VFR." "Four-five."
270.00-299.06	0.095	0.102	0.103	0.117	0.116	"Very stable right now." "We're doin' it—we have a mile now."
300.00-329.06	0.117	0.091	0.171	0.127	0.091	"Four-Eight Golf. We'd like vectors if possible." "Four-Eight Golf." "We got as low as three-five."
330.00-339.06	0.002	0.004	0.003	0.001	0.002	NO VOCALIZATIONS
340.00-369.06	0.098	0.085	0.131	0.118	0.120	"Uh...they vary...about three quarters to a mile visibility...stable." "Yes."

CONTINUED...

TAPE: 48 Golf-Air

Table 5b.

SECONDS INTO TAPE	AUTOCORRELATION PERIODS MEANS FREQUENCY BINS 5-9 Hz 5 Hz	6 Hz	7 Hz	8 Hz	9 Hz	VOCALIZATIONS
370.00-399.06	0.090	0.087	0.107	0.113	0.099	"Four-Eight Golf." "We got ? -uh- scattered rain, light rain. Uh... still visible, about a mile."
400.00-429.06	0.120	0.081	0.109	0.124	0.117	"I'm at three thousand two-hundred right now." "Four-Eight Golf. I doubt if I can make it too."
430.00-439.38	0.005	0.002	0.005	0.002	0.002	"Four-Eight Golf."
440.00-469.06	0.105	0.086	0.130	0.120	0.102	"We're down to three...uh...three thousand fifty-two-five." "One-Eight-Two."
470.00-499.06	0.089	0.071	0.140	0.101	0.113	"O.K. It looks like I can go down this one little small layer. You give me a vector down to the runway?" "Four-Eight Golf."
500.00-529.06	0.130	0.086	0.120	0.139	0.092	"Ninety." "Four-Eight Golf." "It is two-one-zero."
530.00-539.38	0.003	0.002	0.004	0.002	0.002	"Four-Eight Golf."
540.00-569.06	0.092	0.105	0.104	0.130	0.121	"We're at two thousand. It's not very clear at all right here." "That's affirmative."

CONTINUED...

TAPE: 48 Golf-Air

Table 5c.

SECONDS INTO TAPE	AUTOCORRELATION PERIODS MEANS FREQUENCY BINS 5-9 Hz					VOCALIZATIONS
	5 Hz	6 Hz	7 Hz	8 Hz	9 Hz	
570.00-599.06	0.113	0.085	<u>0.166</u>	0.081	0.093	"Four-Eight Golf." "It's *one-eight-zero*."
600.00-629.06	0.088	0.105	<u>0.142</u>	0.126	0.087	"Four-Eight Golf _?_. We kinda' have the strobe lights in sight." "One-two-zero." "One-two-
630.00-639.38	0.002	0.001	<u>0.003</u>	0.002	0.001	zero, roger."
640.00-669.06	0.102	0.125	0.091	0.121	<u>0.127</u>	"We're at two thousand. Can I _?_ down?" "Four-Eight Golf." "Four-Eight Golf. We've got it in sight."
670.00-699.06	0.107	0.119	<u>0.126</u>	0.124	0.079	"That's affirmative!" "Thank you very much."
700.00-729.06	<u>0.136</u>	0.133	0.122	0.109	0.103	"Four-Eight Golf, thank you very much."

THIS PERIOD IS SHARED WITH 1ST VOCALIZATION OF GROUND #1

TAPE: 30 Golf-Ground

TABLE 6. Peak FM analysis for Air Traffic Controller vocalizations. Tape is dubbed to contain only the controller's voice, with pauses between speech segments minimized.

SECONDS INTO TAPE	AUTOCORRELATION PERIODS MEANS FREQUENCY BINS 5-9 Hz: 5 Hz	6 Hz	7 Hz	8 Hz	9 Hz	VOCALIZATIONS
450.00-479.06	<u>0.115</u>	0.113	0.115	This segment shared with Air		"Cessna Three-Zero Golf, roger, say altitude."
480.00-509.06	<u>0.147</u>	0.131	0.089	0.112	0.104	"Cessna Three-Zero Golf, roger. Do you have a transpounder?" "Cessna Three-Zero Golf, roger. Do you have a transpounder?" "Cessna Three-Zero Golf, roger."
510.00-539.06	<u>0.173</u>	0.121	0.114	0.083	0.094	"Do you have a transpounder?" "Three-Zero Golf, say heading." "Cessna Three-Zero Golf, say altitude." "Cessna Three-Zero Golf, San Antonio Approach Control, over."
540.00-569.06	0.030	<u>0.131</u>	0.104	0.101	0.089	"Cessna Three-Zero Golf I hear you, how do you read me?" "O.K. uh...Three-Zero Golf, what is your position in altitude?" "Three-Zero Golf,
570.00-599.06	<u>0.153</u>	0.135	0.109	0.120	0.084	what is your heading?" "Three-Zero Golf, roger. You're uh...on the zero-six-zero radial...northeast of the San Antonio vortex on a heading of due north. Turn left direct to San Antonio vortex, over."

CONTINUED...

TAPE: 30 Golf-Ground

Table 6a.

SECONDS INTO TAPE	AUTOCORRELATION PERIODS MEANS FREQUENCY BINS 5-9 Hz: 5 Hz	6 Hz	7 Hz	8 Hz	9 Hz	VOCALIZATIONS
600.00-629.06	<u>0.180</u>	0.096	0.090	0.092	0.082	"Three-Zero Golf, roger. Can you maintain VFR at one-zero thousand?" "Uh...Three-Zero Golf, roger. What is your present flight conditions ? ?" "Three-Zero Golf, roger. Uh...maintain VFR,
630.00-659.06	0.124	<u>0.143</u>	0.087	0.094	0.110	remain this frequency and advise your intentions." "Roger, and is your call sign uh...Three-Zero Golf or Three-Zero Quebec?" "Cessna Three-Zero Golf, roger. Are you instrument rated and equipped?"
660.00-689.06	<u>0.142</u>	0.130	0.137	0.074	0.089	"Cessna Three-Zero Golf, roger. Continued inbound to the San Antonio VOR, maintain VFR and I'll attempt to get you down when I get you west of the vortex." "Cessna Three-Zero Golf, I'm not paining any weather east of San Antonio at
690.00-719.06	0.107	<u>0.112</u>	0.109	0.103	0.088	the present time." "Three-Zero Golf, San Antonio Approach Control, can you maintain VFR?" "Can you climb and maintain VFR, Cessna Three-Zero Golf?" CONTINUED...

TAPE: 30 Golf-Ground

Table 6b.

SECONDS INTO TAPE	AUTOCORRELATION PERIODS MEANS FREQUENCY BINS 5-9 Hz: 5 Hz	6 Hz	7 Hz	8 Hz	9 Hz	VOCALIZATIONS
720.00-749.06	<u>0.144</u>	0.086	0.119	0.123	0.101	"Three-Zero Golf, roger. I do not have you in radar contact as yet." "Cessna Three-Zero Golf, say your heading." "Cessna Three-Zero Golf, roger. Can you keep your wings low and everything at ten thousand feet and maintain ten?"
750.00-779.06	<u>0.162</u>	0.103	0.115	0.095	0.099	"Cessna Three-Zero Golf, roger, are you...declaring an emergency?" "Affirmative! Three-Zero Golf, roger, request your number of uh...persons on board and uh...what is your full call sign?"
780.00-809.06	0.101	<u>0.112</u>	0.103	0.091	0.094	"Cessna Three-Zero Golf, radar identification turn right heading two-seven-zero." "Cessna Three-Zero Golf, say type of Cessna."
810.00-839.06	<u>0.174</u>	0.114	0.099	0.077	0.073	"Cessna Three-Zero Golf, roger. The target I was observing did not turn. Turn left direct San Antonio vortex and resume normal navigation." "Cessna Three-Zero Golf, San Antonio Approach Control, to what extent do you need our help?"
840.00-869.06	0.157	<u>0.158</u>	0.101	0.107	0.087	"Cessna Three-Zero Golf, key your mike for five seconds." "Cessna Three-Zero Golf, say altitude." "Cessna Three-Zero Golf, roger. Are you

CONTINUED...

TAPE: 30 Golf-Ground

Table 6c.

SECONDS INTO TAPE	AUTOCORRELATION PERIODS MEANS FREQUENCY BINS 5-9 Hz: 5 Hz	6 Hz	7 Hz	8 Hz	9 Hz	VOCALIZATIONS
870.00-899.06	<u>0.128</u>	0.115	0.117	0.077	0.068	still on a heading of two-four-zero?" "Roger, continue heading two-four-zero inbound for San Antonio VOR." "Cessna Three-Zero Golf, roger."
900.00-929.06	0.126	<u>0.130</u>	0.093	0.094	0.086	"Cessna Three-Zero Golf, roger. Cessna Three-Zero Golf, what was your last known position?" "Roger."
930.00-959.06	0.098	0.099	<u>0.131</u>	0.117	0.118	"Cessna Three-Zero Golf, the San Antonio weather is measured ceiling, two thousand three-hundred broken, eight thousand overcast, visibility one-five, altimeter two-niner-niner-one." "Cessna Three-Zero Golf, you turning left at a heading of two-four-zero? Is that correct?"
960.00-989.06	0.130	<u>0.154</u>	0.086	0.105	0.080	"Cessna Three-Zero Golf, for radar identification turn left heading one-eight-zero." "Roger, turn right heading two-seven-zero."
990.00-998.91	0.081	0.133	0.104	0.126	<u>0.143</u>	"Cessna Three-Zero Golf, roger, and heading two-seven-zero?"
0.00-29.06	0.112	<u>0.120</u>	0.095	<u>0.120</u>	0.081	"Cessna Three-Zero Golf, key your transmitter for five minutes." "Uh...correction, make that five seconds."

CONTINUED...

TAPE: 30 Golf-Ground

Table 6d.

SECONDS INTO TAPE	AUTOCORRELATION PERIODS MEANS FREQUENCY BINS 5-9 Hz: 5 Hz	6 Hz	7 Hz	8 Hz	9 Hz	VOCALIZATIONS
30.00-59.06	0.123	<u>0.129</u>	0.120	0.084	0.103	"Uh...five seconds." "Cessna Three-Zero Golf, roger, turn left direct to San Antonio vortex, not to descend below three thousand feet." "Cessna Three-Zero Golf, roger."
60.00-89.06	<u>0.143</u>	0.124	0.090	0.079	0.087	"La Vernia one-one-two-one-zero." "Two-ten from La Vernia?"
90.00-119.06	<u>0.172</u>	0.134	0.118	0.093	0.071	"Three-Zero Golf, keep La Vernia tuned in and rotate your course selector until the needle centers, and then reads TO and then give me that heading." "Zero-two-zero, roger. And that was a TO indication Cessna Three-Zero Golf?"
120.00-149.06	<u>0.144</u>	0.116	0.108	0.079	0.065	"Cessna Three Zero Golf, roger. Tune back into San Antonio VOR and rotate your course selector until the needle centers and reads TO." "Zero-three-five-two San Antone. Is that correct Three-Zero Golf?" "Cessna Three-Zero Golf, what is your heading now?"
150.00-179.06	<u>0.128</u>	0.091	0.094	0.096	0.098	"Cessna Three-Zero Golf, roger." "Three-Zero Golf, roger. Are you in a turn?" "Roger. Cessna Three-Zero Golf, did you fly east of Luling at all?"
180.00-209.06	<u>0.184</u>	0.142	0.103	0.091	0.096	"O.K. and uh...approximately how long did you fly east of Luling before you turned and preceded westbound?" "Roger. Suggest you do not descend below three thousand."

END

DATE
FILMED

9 83

DT