THE PLANAR CORPORATION

PSYCHOLOGICAL STRESS MEASUREMENT

THROUGH VOICE OUTPUT ANALYSIS

Harry J. Older
Larry L. Jenney

March 1975

Distribution of this report is provided in the interest of
information exchange.   Responsibility for the content resides in the
authors and organization that prepared it.

SUMMARY

Audio tape recordings of selected Skylab communications were processed by the Psychological Stress Evaluator (PSE) manufactured by Dektor Counterintelligence and Security, Inc., Springfield, Virginia. Strip-chart tracings were read "blind" and scores were assigned based on characteristics reported by the manufacturer to indicate psychological stress. These scores were analyzed for their empirical relationships with operational variables in Skylab judged to represent varying degrees of situational stress. Although some statistically significant relationships were found, the technique was not judged to be sufficiently predictive to warrant its use in assessing the degree of psychological stress of crew members in future space missions.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

PSYCHOLOGICAL STRESS MEASUREMENT

THROUGH VOICE OUTPUT ANALYSIS

Harry J. Older and Larry L. Jenney

The Planar Corporation

CHAPTER I

BACKGROUND

Problems in the Measurement of Stress

The detection and measurement of psychological stress has been a basic, but somewhat elusive, goal of behavioral science. The ability to determine objectively and quantitatively the internal, psychological state of the individual would have considerable practical value in assessing the capacity to perform work and in predicting situations where performance degradation might occur as a result of environmental factors, workload, task difficulty, or equipment design.

Previous research in stress measurement has usually approached the problem in one of two different ways, both making use of the relationships between the psychological and physiological aspects of behavior. One approach is the analysis of body fluids such as blood, urine, or saliva to determine the presence of hormonal and waste products which have been shown empirically to be associated with the human organism under stress. One is able to deduce, after the fact, that a certain amount of stress has occurred because of the traces which are left in body fluids. The other approach involves measurement of changes in physiological processes which occur during states of presumed stress. The processes most commonly monitored are cardiovascular activity, skin conductivity, respiration,

and electrical activity in the brain. While both methods involve estimation of psychological state by means of measuring related changes in physiological condition, the first is concerned with aftereffects, and the second with concurrent effects. One deals with the bio-chemical products of stress; the other with dynamic processes of the organism while it is under stress.

A considerable amount of work has also been done in which changes in performance on the subject's normal work task or on artificial tasks have been considered as indication of stress and/or fatigue.

Some successes have been achieved in stress measurement, but most techniques are still unsatisfactory on one or more of the following grounds.

o They are intrusive or interfering. Subjects are aware that they are being measured. This often induces either an extraneous source of stress or an artificiality in the test situation. Taking measurements on a subject while he is engaged in an operational task may also interfere with task performance or restrict normal activity.

o Some stress measures, especially those which make use of performance as an index of stress, suffer from being too specific to the system or task for which they were developed. Generalization beyond the particular test situation or the specific system under study is usually difficult to substantiate or is subject to great inaccuracies.

o For convenience of measurement, it is often necessary to conduct stress studies in a simulated operational situation. The element on non-realism and the synthetic nature of the operational setting act to diminish stress indications or to obscure the true effect of stress in "real world" circumstances.

o The attachment of sensors to individuals or the interruption of

normal routines to take special stress measures is as cumbersome
as it is obtrusive and interfering. The equipment for obtaining
and storing body fluid samples, the electrodes and pressure cuffs
which must be fitted to subjects, and the apparatus for recording
experimental data are just a few of the impedimenta associated with
conventional stress measures.

The foregoing suggests that present techniques for detecting and mea-
suring psychological stress fall short on either methodological or practi-
cal grounds. The measures which can be applied in operational circumstances
usually do not provide a clear enough picture of the individual's psycho-
physiological state. On the other hand, those measures which do discrim-
inate sensitively along psychophysiological dimensions are usually not
applicable in operational contexts because they are interfering, cumber-
some, or otherwise impractical.

There is a need for a psychological stress measure which is non-obtru-
sive and which can be applied in actual operational circumstances without
interference with performance routines. Ideally, the technique should
produce objective and quantifiable stress indices. Also, it should be
simple to apply, and the need for ancillary data collection and processing
apparatus should be minimal. The technique should yield measurements which
are repeatable and reliable, in that they are consistent across individuals,
test situations, and experimenters.

The development of such a measure of psychological stress would have
many advantages. As a basic research tool, a measure which could be applied
to individuals while they were performing operational tasks in actual op-
erational settings would provide a much more detailed and realistic under-
standing of stress and human performance capability. The practical appli-
cations of a tool which would permit in situ monitoring of psychophysio-
logical state or performance capacity are many.

Voice Analysis as a Possible Measure of Stress

Previous research in the field of voice analysis has indicated that the psychophysiological state of the speaker may manifest itself in a variety of ways in the acoustic domain -- rate and timing of voice production, pitch and volume of the voice signal, vocal articulation, and arrythmia, to name a few*

Generally, the human voice mechanism produces two types of sound--the fundamental frequency and the formant frequencies. The fundamental frequency is a product of the vocal cords which vibrate when expelled air is forced through the partially closed glottis. The vibrations of the vocal cords, which provide most of the acoustic power for speech, vary between 60 and 350 Hz, depending upon the age and sex of the speaker and the intonation applied.

The second type of sound, the formant frequencies, result from resonance of the cavities in the head (throat, mouth, nose, and sinuses) when excited by sound of the fundamental voice frequency. The formants range generally from 500 to 4500 Hz and appear in distinct frequency bands which correspond to the resonant frequency of the individual cavities. The formant wave forms are ringing signals, as opposed to the rapid decay signals of the fundamental voice frequency. When voiced sounds are uttered, the wave forms of the fundamental voice frequency are imposed upon the formants as amplitude modulation.

Physiologically, the formant frequencies are determined by the characteristics of the head cavities. The fundamental frequency is primarily

---

*The research literature on voice characteristics as indicators of stress is rather sparse. The short bibliography presented in Appendix B , while not exhaustive, covers the most significant research in the field in recent years. The consensus of the investigations is that voice output and psychophysiological state are related, but there is uncertainty about the precise nature of the relationship and disagreement about which speech characteristics provide the clearest indices of stress.

controlled by the activity of the laryngeal musculature and by the dynamics of the respiratory system (particularly subglottal pressure). The fundamental frequency is also directly influenced by the degree of organization and coordination between and within these physiological processes. It is, therefore, possible that wave form characteristics of the fundamental voice frequency may exhibit patterns which are associated with the psychophysiological state of the speaker. Under conditions of systemic disruption (due, for example, to drugs or alcohol) or under conditions of psychological stress, a certain amount of neuromuscular disorganization or impairment can be expected in the laryngeal and respiratory functions. The fundamental voice frequency, which is directly influenced by both these physiological mechanisms, may therefore reflect these conditions through changes in its signal characteristics. (Grether, 1971)

This conclusion is, in part, based on the findings of other investigators who have examined the relationship between psychophysiological state and voice characteristics. For example, Huttar (1968) observed that changes in laryngeal configuration can generally be attributed to an increase in laryngeal tension and muscular activity. These increases in muscular activity can in turn be attributed to the increase in muscular tension throughout the body, which appears to be a concomitant of emotion.

Williams and Stevens (1969) have offered a more detailed hypothesis as to the relation between stress and speech characteristics. Their research indicated that, of all the parameters of speech, fundamental frequency exhibited the highest correlation with emotional state. Further, they found that the fundamental frequency could undergo variations which might not be intended by the speaker or be under his overt control, thus providing an index of the speaker's psychophysiological state. Williams and Stevens also noted that the muscular activity in the larynx and the condition of the vocal cords were likely to have a more direct effect on the sound output (and, in particular, on the fundamental frequency) than changes in muscle activity in other parts of the speech-generating system. The reason is that the vocal cords themselves constitute the primary

5

sound-generating components, whereas other muscles and vocal tract components simply shape the resonant cavities for sound that originates at the vocal cords. Thus, they concluded that analysis of the portion of the speech signal which reflects vocal cord activity (i.e., the fundamental frequency) was more likely to reveal changes brought about by the psychophysiological state of the speaker.

If this is the case, the wave form characteristics of the fundamental voice frequency may be capable of analysis so as to provide a measure of stress which meets the criteria outlined above. This is particularly true of situations where the normal activities of job incumbents require a substantial amount of voice communication, as is the case with astronauts. The availability of magnetic tape recordings of such communication allows retrospective analysis for research purposes and, if such research yields promising results, for possible application in future manned space missions.

The designers of the equipment used in the present study based their approach to the analysis of the fundamental frequency on findings such as those cited above and upon further considerations relating to a phenomenon known as micro-muscular tremor.

The vocal cords and the walls of the major formant-producing cavities are soft tissue immediately responsive to the complex array of muscles which control them. The muscles controlling the vocal cords create both the purposeful and involuntary production of voiced sound and variation of voice pitch. Similarly the muscles controlling the throat, lips, and tongue produce the purposeful and involuntary variation of first formant frequencies. During normal speech, these muscles are performing at only a small fraction of their work capacity. For this reason, in spite of their being employed to change the position of the vocal apparatus, the muscles remain in a relatively relaxed state. During this relaxed state it is thought that the muscles exhibit the minute undulations which normally accompany the activity of any voluntary muscle. These oscillations, known as physiological tremor or micro-muscular tremor, occur at a rate

6

of 8-14 Hz. These tremors appear to be the result of central nervous system activity, although the precise nature of the controlling mechanism is not fully understood at this time.

The micro-muscular tremor manifested in the larynx causes the tension of the vocal cords to vary slightly. These variations produce audibly indiscernible fluctuations in the fundamental pitch frequency of the voice. These shifts about a central frequency constitute a frequency modulation of the fundamental voice frequency. Thus, in normal speech by a person not under stress, there is an inaudible oscillation of the fundamental frequency through a range of 8-14 Hz. For example, for a person whose fundamental voice frequency is 150 Hz, there would be a normal fluctuation of this frequency between roughly 145 and 155 Hz.

When the individual is subjected to moderate psychological stress, the action of the autonomic nervous system is thought to increase muscular tension throughout the body, including the musculature of the larynx. This tension, imperceptible to the individual, is sufficient to suppress the normal micro-muscular tremor in the laryngeal apparatus and thereby to diminish the oscillations of the fundamental voice frequency found in an unstressed individual. As stress increases and the autonomic nervous system gains dominance over central nervous system activity, the micro-muscular tremor is reduced or may disappear altogether. In the voice this is manifested by elimination of the 8-14 Hz frequency modulation of the carrier wave of the fundamental voice frequency. The suppression of this frequency modulation under stress is involuntary in the speaker and inaudible to the listener. However, through appropriate analysis of the voice spectrum, the phenomena can be identified and charted to produce a visual record of these changes in voice characteristics. The theory holds that these changes are related to the psychophysiological state of the speaker at the time he made the utterance.

The Psychological Stress Evaluator (PSE)

The method employed in this study utilized the Dektor PSE device which processes and makes a graphic record of signals produced by the human voice. This device was specifically designed to emphasize those voice characteristics which are indicative of a stress situation and to deemphasize other voice characteristics unrelated to stress. The device is most sensitive to the frequency range associated with the fundamental voice frequency, and it is designed to detect and analyze the 8-14 Hz frequency modulation imposed on the fundamental voice frequency by micro-muscular tremor.

This device, which consists of a signal analyzer and a strip chart pen recorder, is normally used in conjunction with a conventional tape recorder. Voice signals are initially recorded on magnetic tape, then processed through the analyzer circuits, and recorded on a strip chart for subsequent visual analysis and interpretation.

The subaudible effects on the voice thought to be influenced by stress are emphasized in the PSE by means of a combination of amplitude demodulation and selective frequency filtering. The amplitude modulation of the format frequencies (imposed by the frequency modulation of the fundamental voice frequency due to micro-muscular tremor) is detected by the demodulation processes. The frequency filtering process allows the low frequencies associated with micro-muscular tremor effects to pass through the instrument to the pen recorder while attenuating higher frequencies which have no direct relationship to stress. (See Appendix A for a more detailed description of the PSE).

Interpretation of the strip chart tracing (the final output) is accomplished by visual examination of the average level of the recorded signal for specific types of changes. For example, random changes in the output signal are said to indicate a low stress level, a slowly increasing average level to indicate moderate stress, and a steady average level to indicate even higher stress. Certain other characteristics of the output signal

are said to be associated with level of stress, the most notable of these
being cyclic rate change (not found in this sample) and amplitude suppres-
sion. Figure 1 contains examples of the output signal wave forms which
were scored for analysis in this study.

The output signal characteristics upon which the interpretation is
based are inferred from the physiological theory underlying the technique
and from the design of the PSE electronics. The manufacturers of the de-
vice have reported in several documents that these characteristics are
stress indices. However, since the PSE was developed primarily as
an aid to interrogation and criminal investigation, any empirical evidence
of its validity relates to applications of the techniques to detect stress
due to willful deception and guilty knowledge. Thus, the more general use
of the device to measure psychological stress attributable to factors such
as workload, fatigue, and emotional factors has not yet been investigated
in any extensive and systematic way.

Since the present investigation is completely unrelated to its use in
interrogation, no attempt will be made here to review the very sparse liter-
ature concerning that application.

It should be clearly understood that the use of the PSE in the present
study involved procedures and techniques radically different from those in
which the device is normally employed. The typical use of the device is
in a situation similar to that in which the Polygraph is used, i.e., a
structured interview situation. Typically, a protocol for the interview
is constructed in advance, the interviewee is well aware of the nature and
purpose of the interview, and the interviewer is a trained interrogator.
The purpose is most often the detection of stress as an aid to investiga-
tion of the possible involvement of the interviewee in matters where the
truth or falsity of his responses is at issue. No such condition exists
in the material dealt with in this project. The samples of communication
included in this study are, in every case, the ordinary work-related con-
versation between astronauts and ground personnel. The situation is so

9

FIGURE 1

SIGNAL WAVE FORM PATTERNS
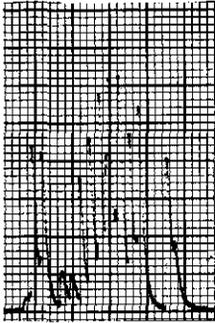
RANDOM
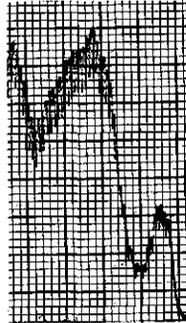
DIAGONAL
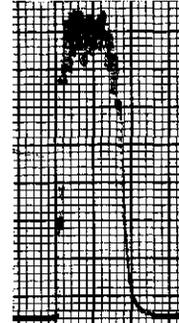
BLOCKING
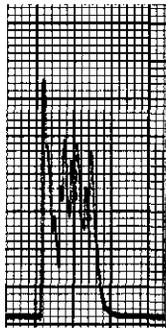
LEADING EDGE

AMPLITUDE SUPPRESSION

different from that in which the PSE is normally used that the findings of this study should be considered completely irrelevant to the purpose for which the device was developed. No inference should be drawn concerning the usefulness of the technique in any application beyond that investigated here.

Rationale for the Selection of Skylab
Missions for Evaluation of Voice Analysis

Preliminary efforts in the project included examining communications material from all NASA manned space programs. The Apollo program was studied in detail to find potentially suitable material to be utilized in the evaluation of voice analysis techniques as measures of psychological stress. Since the emphasis in the evaluation was on psychological stress many of the most interesting phases of Mercury, Gemini, and Apollo were felt to be inappropriate since they included physical stress, e.g. unusual gravitational forces, physical work, etc. Skylab missions, on the other hand, are of sufficient duration to allow longitudinal study of crew performance over time; and they contain numerous highly technical tasks which are carried out repetitively. This permits an assessment of stress as a function of task difficulty. Skylab missions III and IV are particularly appropriate because they are less contaminated by physical stress and irregular events than was the case with Skylab II where there was considerable heat stress in the early part of the mission and where maintenance and repair of the space station disrupted the normal schedule of activities. Consequently all communications materials for use in this project were drawn from Skylab missions III and IV. (Certain preliminary analyses to determine rater reliability and to refine scoring methods made use of communications from Apollo missions).

CHAPTER II

METHODOLOGY

Introduction

The fundamental hypothesis underlying this study is:

> An individual's current psychophysical
> state is accompanied by a change in the
> fundamental frequency of the voice, and
> analysis of fundamental frequency changes
> will reveal signal characteristics which
> can serve as related and accurate measures
> of stress of the speaker.

It was further postulated that, if such a technique were to be useful
to NASA as a measure of psychological stress it should have enough sensi-
tivity to permit discrimination among degrees and kinds of situational
stress of concern in future manned missions, e.g., shift length, workload,
length of mission, and type of activity. In other words, sophisticated
analytical tools are not required to reveal that situations of great danger
or those requiring extreme physical effort are stressful. What is needed
is information concerning the differential stress-producing effects or
attributes of the situation which are subject to the control of system
designers, mission planners, and crew members (in space and on the ground).
Thus, it was felt necessary to obtain a sample of communications which rep-
resented a normal range of such situations but which avoided extremes of
psychological or physiological stress.

For those not familiar with the Skylab program, the following brief
general background is given.

The Skylab Program was established for four purposes: (a) to deter-
mine man's ability to live and work in space for extended periods, (b) to
extend the science of solar astronomy beyond the limits of Earth-based
observations, (c) to develop improved techniques for surveying Earth

12

resources from space, and (d) to increase man's knowledge in a variety of other scientific and technological regimes.

Skylab, the first space system launched by the United States specifically as a manned orbital research facility, provided a laboratory with features not available anywhere on Earth. These included: a constant zero gravity environment, Sun and space observation from above the Earth's atmosphere, and a broad view of the Earth's surface.

Principal scientific and technical objectives of the program included:

Obtaining data for evaluating crew mobility
and work capability in both intravehicular
and extravehicular activity,

Obtaining medical data on the crew for use
in extending the duration of manned space
flights,

Obtaining medical data for determining the
effects on the crew which result from a
space flight of up to 89 days duration,

Obtaining solar astronomy data for con-
tinuing and extending solar studies beyond
the limits of Earth-based observations (ATM),

Obtaining data on the comet Kohoutek beyond
the limits of Earth-based observation,

Performing assigned scientific, engineering,
and technological experiments.

Skylab III was launched on July 28, 1973 and splashed down 59 days, 11 hours, and 9 minutes later. Skylab IV was launched on November 16, 1973 and returned to earth after 84 days, 1 hour and 17 minutes.

While the general configuration of each mission was developed in advance of the launch, specific activities for each day were designed to take advantage of unique conditions or opportunities. For example, forecasts of cloud-free EREP sites and ground observatory predictions of unusual solar activity had a bearing upon when EREP passes and ATM runs were

scheduled. The normal Skylab crew workday started at 6 a.m. and ran until 10 p.m. (Houston time).

During each mission the astronauts operated and monitored about 60 items of experimental equipment and performed a wide variety of tasks associated with the several hundred Skylab scientific and technical investigations. Depending upon experiment scheduling requirements, Skylab crews had a day off about every seventh day.

About two 15-minute personal hygiene periods were scheduled each day for each crewman and one hour and 30 minutes for physical exercise. Additionally, an hour a day was usually set aside for rest and relaxation.

Radio communications were maintained with Mission Control through direct air to-ground radio link whenever conditions permitted. When the spacecraft was not able to transmit directly, an on-board tape recorder was utilized. The material placed on tape was "dumped" or transmitted to a ground station at the first opportunity. Both direct air-to-ground and "dumps" were recorded on magnetic tape on the ground. This library of tapes served as the source of material for this study.

Figure 2 illustrates a typical day in a Skylab mission.

Study Design

Since the Skylab program was not designed for the convenience of this study, it was necessary to make use of independent variables which happened to be available in the normal course of the missions. Thus, some comparisons which would have been very useful were simply not possible. There were, however, several conditions or situations within the missions which make possible a fair evaluation of the usefulness of the voice analysis technique. The study design contained seven independent variables:

      a. Mission (Skylab III vs. Skylab IV)

      b. Mission Day (time into mission)

TYPICAL CREW DAY

| GMT | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 1 | 2 | 3 | 11 |

CDR — ATM | PH | EAT | EREP PASS 1 | PT | PH | ATM | S/HK | * | PRE | | | SLEEP

CMN — SPT — POST SLEEP ACTIVITIES | TVSU | S/HK | M131-1 SUB | PH | ATM | EAT | ETC | M131-1 OBS | M092/M171-1 SUB | PH | ATM | SLEEP | ATM | ACTIVITIES | SLEEP

PLT — EREP TAPE | S/HK | PH | M131-1 OBS | EAT | EREP PASS-1 | M131-1 SUB | M092/M171-1 OBS | PT | PH | S/HK | * | ATM | ACTIVITIES | SLEEP

DAY/NIGHT
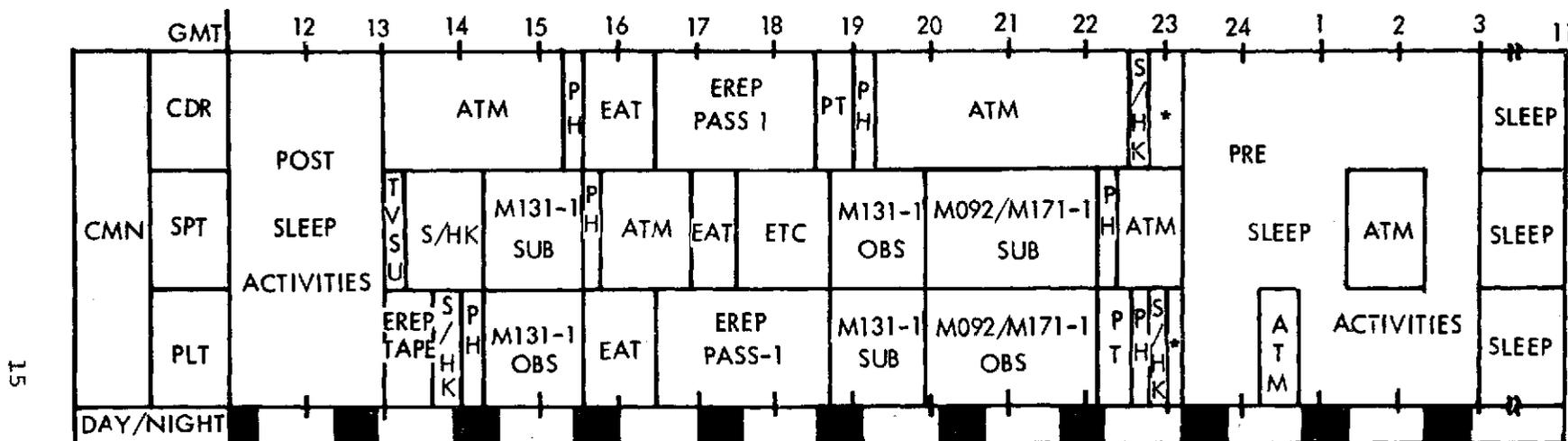
POST SLEEP ACTIVITIES

SYSTEM CONFIGURATION
PH
URINE SAMPLING
T003 EXPERIMENT
BODY MASS MEASUREMENT
BREAKFAST
DINNER PREP
PRD READOUTS
LOAD FILM
REVIEW PADS
STATUS REPORT

S/HK – SYSTEM HOUSEKEEPING

PH – PERSONAL HYGIENE

PT – EXERCISE

TVSU – TV SETUP

* TIME AVAILABLE FOR
  COROLLARY EXPTS

PRE-SLEEP ACTIVITIES

EVENING MEAL
ATM (1 to 2 PASSES)
MISSION PLANNING
RECREATIONAL ACTIVITIES
CONDENSATE DUMP
TRASH AIRLOCK DUMP
FOOD RESIDUE WEIGHING
STATUS REPORT
T003 EXPERIMENT
SYSTEM CONFIGURATION FOR SLEEP
PH
BREAKFAST PREP

FIGURE 2

    c.   Time on Duty (time since wake-up)

    d.   Task (EREP vs. ATM)

    e.   Type of Activity (Task performance vs.
        reporting)

    f.   Crew Position (Commander, Science Pilot, and Pilot)

    g.   Speaker (individual)

Each of these variables is discussed below.

Mission - The two missions studied were the last two manned Skylab missions, III and IV. The first manned mission (Skylab II) was not included since unusual equipment problems resulted in a substantially different environment, physical workload, and other conditions from those in the other two missions. Thus, any differences which might have emerged would not be clearly psychological in nature.

The primary rationale for comparing Missions III and IV stemmed from clear differences between the two missions with respect to schedules and crew attitudes toward workloads. The following quotation from the official NASA report "Skylab Mission Report, Third Visit (JSC-08963)" (Mission IV) indicates the nature of the scheduling problem.

> "... it became apparent during the early part of the third visit that the crew was being over-scheduled relative to the pace to which the crew felt attuned for their longer-duration visit. (Ground personnel later learned that the crew had always intended to work at a somewhat reduced pace, but this fact had not been sufficiently communicated to all concerned.) By the time the first one-third of the visit was over, the ground planners had achieved a better understanding of the desired pace, and adjustments were made to reflect more realistic goals. These reductions were apportioned among the experiment disciplines on the basis of priority and other considerations..."

An examination of the transcripts of communications between the crew and ground personnel indicated that a certain amount of psychological stress attended the resolution of this problem.

Mission III, on the other hand, resulted in no scheduling difficulties and no situation which was thought to have introduced psychological stress. A NASA Mission Report on this mission (MR-14) stated:

> "...12-hour workdays were no problem, and the crew became so proficient that they asked for and were given additional assignments. As a result, the crew completed about 1 1/2 times the work originally planned for them despite a severe bout with motion sickness that hampered them during their first few days in space."

Mission Day (time into mission). The primary concern here was the possible impact of cumulative effects of "stress" over time. Although there was no substantial evidence of long-term effects of fatigue or stress in physiological data or in subjective reports, preliminary data on error rates in the operation of the ATM showed systematic changes. In addition, there was reason to believe that a significant change in the psychological adjustment of the crew of Skylab IV took place following the resolution of the previously discussed scheduling problem. Thus, this variable was included.

The following days were selected for study from each mission.

Skylab III Mission Days	Skylab IV Mission Days
18	18
25	25
32	32
39	39
46	45
53	46
	47
	53
	60
	67
	74

It can be seen that for each mission a sample of communications was taken on a seven day cycle, excluding start-up and wind-down periods. For Skylab IV two extra days were included for the following reason. The scheduling difficulty was resolved on Mission Day 46. Since this represented a period of potential interest from a psychological point of view, the day before and the day after this were included for special study.

Time on Duty. This variable was included to permit analysis of the effects of "time on shift." The sample could not be controlled for this variable, but the selection process resulted in a more or less balanced sampling throughout the work day. Since the time of each activity in relation to the last rest period was known, it was possible to examine the total pool of items for variation along this dimension.

Task (EREP vs. ATM). The two major activities of concern to this project were the conduct of Solar observations using the Apollo Telescope Mount (ATM) and remote sensing operations for studying Earth resources utilizing the Earth Resources Experiment Package (EREP).

The ATM carries an array of telescope packages to permit simultaneous viewing of solar activity in different wavelengths. In addition, it includes the necessary navigational and guidance systems to control attitude and telescope alignment. The consoles for operation of this equipment and associated thermal conditioning and electrical power systems are quite complex and present a demanding task for the astronaut.

The EREP experiments were designed to test and validate remote sensing techniques over a wide spectral region from orbital altitudes. Experiments in EREP permitted simultaneous remote sensing of ground test sites in the visible, infrared, and microwave spectral regions. While the equipment and procedures involved in EREP were not simple, they were, in general, less complex and demanding than those in ATM. Thus, the inclusion of this variable permitted an evaluation of the effects of task difficulty.

For each mission day it was planned to include a sample of communications from each crew member discussing two ATM or EREP operations or one of each. This was possible except for a very small number of days (three). In these cases another scientific experiment of a similar nature was substituted.

Type of Activity (Performance vs. Reporting). In some instances, transmissions were conducted during the actual performance of the ATM and EREP tasks. In other cases, the task was completed, notes made, and the results reported later. This offered an opportunity to study possible differences as a function of the immediacy of the situation. The supposition was that the astronaut would be under more stress during actual task performance than during the post-activity debriefings.

In the final sample, it was possible to counterbalance performance and reporting. For each day, one communications sample was included for "performance" and one for "reporting" for each crew member.

Crew Position (CDR, SPT, PLT). On each mission the three crew members had differing responsibilities. The Commander (CDR), Science Pilot (SPT), and Pilot (PLT) in each mission had, however, for the sample activities chosen, essentially identical responsibilities, namely the conduct of the specific ATM or EREP experiment being carried out. However, it was considered possible that differences would emerge as a function of overall mission responsibilities. Thus, data were maintained separately to permit such comparisons.

Speaker (Individual). This is, of course, merely a further categorization of the above variable. It is clearly very important to identify the contribution made by individual differences among speakers to any comparison.

Thus, the basic design of the study was as shown in Table 1.

SKYLAB III   DURATION

CREW ACTIVITY		14 Aug	21 Aug	28 Aug	4 Sep	11 Sep	18 Sep	DATE
		226	233	240	247	254	261	DAY of YEAR
		18	25	32	39	46	53	Mission Day
CDR	Perf.	ATM	ATM	ATM	EREP	EREP	EREP	
	Rept.	ATM	ATM	EREP	ATM	EREP	EREP	
SPT	Perf.	ATM	(1)	ATM	EREP	EREP	(2)	
	Rept.	ATM	ATM	ATM	ATM	ATM	ATM	
PLT	Perf.	ATM	ATM	ATM	ATM	ATM	ATM	
	Rept.	ATM	ATM	ATM	EREP	ATM	ATM	

SKYLAB IV

CREW ACTIVITY		3 Dec	10 Dec	17 Dec	24 Dec	30 Dec	31 Dec	1 Jan	7 Jan	14 Jan	21 Jan	28 Jan	DATE
		337	344	351	358	364	365	001	007	014	021	028	DAY of YEAR
		18	25	32	39	45	46	47	53	60	67	74	MISSION DAY
CDR	Perf.	EREP	ATM	ATM	ATM	ATM	ATM	EREP	ATM	ATM	EREP	EREP	
	Rept.	ATM	ATM	ATM	ATM	ATM	ATM	EREP	ATM	ATM	EREP	ATM	
SPT	Perf.	ATM	ATM	ATM	ATM	ATM	ATM	ATM	EREP	ATM	ATM	ATM	
	Rept.	ATM	EREP	EREP	ATM	ATM	ATM	ATM	ATM	ATM	ATM	EREP	
PLT	Perf.	EREP	EREP	EREP	ATM	EREP	ATM	EREP	EREP	EREP	EREP	ATM	
	Rept.	ATM	EREP	ATM	ATM	ATM	EREP	ATM	EREP	ATM	EREP	(3)	

(1) Rate gyro temperature test   (2) Video Tape Recorder   (3) Photographic log

TABLE 1 - Study Design

Communication Samples. For each cell in Table 1 a tape recording was obtained. Through this report these will be referred to as "episodes." From each episode 20 segments of communications were chosen for study. Each segment consisted of a statement made by the crew member. These were typically two to four words in duration and were almost invariably substantive in nature. For example, if the crew member made the following statement, "Well, we're coming up on the West coast of Greenland now and I can see a large area of dark blue water surrounded by much lighter green--the icebergs are all in the blue water," the phrases "West coast of Greenland" and "large area of dark blue water" would be candidates for inclusion in the sample.

Words or phrases uttered with special emphasis, e.g. "STOP" or "MARK!" were avoided. Words or phrases serving as filler, e.g. "don't you know," or emphasis "Well, how about that!" were also not included. Table 2 is an illustrative transcript from one of the missions and indicates the types of material included in the study. In this case, the Pilot (PLT) is performing an EREP experiment. Brackets and underlining illustrate the types of phrases charted and included in the analysis.

Thus, the study design included seven independent variables among which major comparisons can be made. The basic matrix resulted in 102 cells (3 crewmen x 2 types of activity x 6 days for Skylab III = 36, plus 3 crewmen x 2 types of activity x 11 days for Skylab IV = 66). Each cell contains 20 statements, yielding a total item pool of 2,040.

Secondary independent variables (those which might have had an effect on voice tracings but which were not of useful significance as stress indicators) were:

 o Voice quality (the technical quality of the recording, e.g., noise),

 o Recording mechanism (air-to-ground vs. tape-dump).

21

TIME	SPEAKER	MESSAGE
17 14 30	PLT	MARK, [SHUTTER SPEED to MEDIUM.]
	CDR	Okay, whistling over the coast of Florida.
	CC	PLT, Houston. While we got a short break here, I would like to advise you we have seen FILM ADVANCE MALF lights before on mags that have not been used previously, and so we don't think that's anything unusual.
	PLT	Okay. [I cycled the POWER, OFF] after the last 190 sequence and put it back on. And then when I did my sequence this time, it's in a sequence right now, I only have [a 5 light right now.]
	CC	Okay.
	PLT	192, [MODE to STANDBY at 15:20.]
17 15 20	PLT	MARK, STANDBY. [192 POWER, OFF,] and waiting for 16:30
	PLT	[192 POWER, OFF], okay.
	CDR	Okay. Got the nadir swath going, the weather over Florida is beautiful. I should say it was beautiful.
	CDR	Starting to pick up clouds now. Nice blue water.
	PLT	Okay, Ed, at 16:30, ETC, STANDBY.
	PLT	Still have [an ALTIMETER UNLOCK light] but your [READY light is remaining on].
	CDR	Still scattered clouds.
17 16 54	SPT	MARK. ALTIMETER, STANDBY and MODE, 5. Stand by for 17:13.
	CDR	Lots of cloud street between ─ ─
	PLT	Stand by.
	CDR	─ ─ 16 minutes and ─ ─
17 17 13	PLT	MARK. [17:13, ALTIMETER, ON.]

TABLE 2 ─ Sample of Communications Material

Voice Analysis Procedures

For each of the 102 cells noted above, typewritten transcripts of communications for the appropriate day were obtained. These were examined to identify appropriate task performance or reporting. For selected episodes, tape recordings were obtained from Johnson Space Center. These recordings were made either directly from air-to-ground (A/G) transmissions or from delayed (Dump) transmissions. The PSE manufacturers indicate that some differences can be expected in tracings due to the fact that an additional "generation" is involved in the dump tapes. Empirical analyses by the present investigators indicate that, while detailed differences resulted from successive re-recordings, the essential nature of the resultant tracing patterns did not appear to be significantly affected.

Tapes received fron NASA were recorded at 4.7 cm (1 7/8 in) per second. In order to make these suitable for use with the PSE, these were re-recorded at 19 cm (7 1/2 in) per second. Since this was done for both A/G and dump tapes any generation changes were constant.

All tape reproductions and all PSE charting was done with a UHER tape recorder, Model 4000 Report 1C.

PSE charting of all episodes was done by the same person, trained by and following techniques recommended by the equipment manufacturer.

To insure that chart interpretation would be as free as possible from bias due to knowledge of the episode, all sooring was done "blind." The chart for each episode was identified for the scorer only by a randomly assigned number. Thus, scoring was based exclusively on interpretation of the voice trace patterns which constituted the dependent variables of the analysis, and the scorer had no information concerning the nature of the episode itself, i.e., the independent variables of mission, task, time of day, and speaker.

For each episode, twenty utterances (individual words or short phrases) were charted. These were presented in one continuous strip chart and analyzed as a group. As mentioned above, each chart (episode) was assigned an identification number, and the utterances within each were numbered serially in the order in which they occurred.

Scoring Procedure

Step 1. The entire chart (20 segments) was examined and an overall rating assigned for the amount of patterning present. This score was labeled "Overall Patterning Estimate." These, and all, other ratings on "scores" utilized a scale of 5; 1 = no patterning--5 = high incidence of patterning.

Step 2. The entire chart was again examined--at a later time--and an overall estimate assigned on each of the dependent variables of concern, e.g.
- o Blocking Pattern
- o Diagonal Pattern
- o Leading Edge
- o Amplitude Suppression

Step 3. Again at a later date, each chart was examined in detail and a specific score was assigned for each of the 20 segments. The mean was then calculated for the 20 scores assigned to each variable. This mean became the "Score" for each variable in each cell of the analysis.

Step 4. A "Total Score" was developed by calculating the mean of the scores for the four variables listed above.

Specific Scoring Rules

Blocking Pattern
- a. Assign score of 1 to 5 to entire sample of twenty utterances based on overall estimate of the degree of blocking pattern throughout. (1 = no blocking, 5 = high incidence of blocking).

b. Examine each speech element in the utterance for the blocking pattern using a scale of 1 to 5. (1 = no blocking, 5 = maximum blocking).

c. Assign the highest score obtained for an element as the score for the utterance.

Diagonal Pattern

Follow the same procedure as for Blocking Pattern, using a scale of 1 to 5 (1 = no diagonal, 5 = maximum diagonal).

Leading Edge

Follow the same procedure as for Blocking Pattern, using a scale of 1 to 5 (1 = irregular, sloping leading edge, 5 = straight, perpendicular leading edge).

Amplitude Suppression

Follow the same procedure as for Blocking Pattern, using a scale of 1 to 5 (1 = little or no variability of amplitude within speech elements, 5 = extreme variability).

All scoring was done by a research scientist who had completed the Dektor training program and, who had participated in several previous analyses of tracings in other applications of a similar nature. Thus, problems of inter-judge reliability were avoided. Preliminary studies, using Apollo communications showed that inter-judge reliability was acceptable (ranging from .70 to .92 on scores assigned to individual variables, N = 253). No satisfactory method for determining intra-judge reliability was deemed suitable without allowing a time period between scorings which exceeded that available.

In addition to the scores on the 102 episodes assigned by Planar personnel in the manner described above, 48 episodes were scored by the Chief Instructor of Dektor Counterintelligence and Security, Inc. Almost all of his experience in analyzing voice tracings has been in applications invol-

ving structured interview situations. Thus he was not accustomed to a pro-
cedure which involved "blind" scoring of charts where the speaker, the
situation, and the content are unknown. On the other hand, he is clearly
much more experienced than Planar personnel in analyzing tracings of voice
communications as produced by the PSE. The scores assigned by the repre-
sentative of Dektor are referred to in the Results section as "Dektor Totals."

Dependent Variables

Thus, the primary dependent variables of interest were:

o Planar Total (the arithmetic mean of the following four
    scores)
    - Blocking score
    - Diagonal score
    - Leading Edge score
    - Amplitude Suppression score
o Dektor Total score.

Since detailed scoring of the voice tracing charts was an extremely
tedious task, it was felt to be important to determine the incremental
value of such detailed scoring over that which would result from a pro-
cedure which involved a brief review of the entire chart and the assign-
ment of an estimated score for overall patterning and for each of the
specific types of patterns (See steps 1 and 2 on p. 24 above). This
procedure resulted in five additional scores:

o Total Estimate
o Blocking Estimate
o Diagonal Estimate
o Leading Edge Estimate
o Amplitude Suppression Estimate.

A separate correlational analysis was performed on these scores and
is reported in the following chapter.

CHAPTER III

RESULTS

The essential question in this study was, "Are there non-chance relationships among the operational (independent) variables in the Skylab missions and the characteristics of the selected voice communications of sufficient magnitude and consistency to be of use to NASA in future manned missions?"

In order to determine the answer to this basic question, several analyses were performed. The most general of these involved tests of the significance of the differences between voice tracing scores for various sub-groups on each of the independent variables. For example, the scores for all episodes in Mission III were pooled and compared with the similar pooled values from Mission IV. Where the independent variable fell naturally into two categories, such as the Mission III vs. Mission IV situation, the basic comparison was a "t" test of the significance of the difference between means. Where more than two categories were concerned, i.e., a determination of the effect of speaker on scores, an analysis of variance was performed. Table 3 below summarizes the principal findings of these analyses.

Table 3 cross-tabulates independent and dependent variables, indicating those situations where significant relationships were found. It can be seen that most operationally important comparisons yielded results which did not meet statistical significance standards.

There were, of course, several comparisons which, when total scores were compared, resulted in statistically significant differences. The practical usefulness of the differences, however, is another matter. In order to illustrate the problems which would be involved in attempting to use a technique with such weak relationships to the operational variables,

TABLE 3

SUMMARY STATISTICS

SCORES	Mission*	Mission Day*	Time of Day**	Task*	Activity*	Crew Position**	Speaker**	Cells
Blocking		$p<.001$ (2)						102
Diagonal		$p<.001$ (2)						102
Leading Edge		$p<.001$ (2)						102
Amplitude Suppression	$p<.001$ (1)	$p<.001$ (2)			$p<.05$ (3)			102
Planar Total Score		$p<.001$ (2)						102
Dektor Total Score	$p<.05$ (1)					$p<.01$ (4)	$p<.001$	48

(1) Mission IV scores higher than Mission III scores for Planar Total, reverse for Dektor Total.
(2) Scores lower in later portion of Mission IV.
(3) Scores higher for reporting than for performing (opposite prediction).
(4) PLT scores highest, SPT scores lowest.
 *   t test
 ** ANOVA

the results for each independent variable are discussed separately below.

Unless otherwise indicated, the scores in all tables are the arithmetic means derived from individual scores assigned to the 20 voice messages for each episode. In all cases the number of episodes on which the means are based is 102, except for the Dektor Total Scores which are based on 48 episodes.

Mission

The primary comparison here was between Mission III and IV. It will be remembered that Mission IV was judged by most NASA personnel to be somewhat more stressful than Mission III since it was longer and some misunderstandings developed between the crew and ground personnel (see p. 16 above).

Table 4 below presents the mean scores for the two missions and relevant statistics for each comparison.

TABLE 4

COMPARISON OF SCORES FOR THE TWO MISSIONS

Mission	Blocking	Diagonal	Leading Edge	Amplitude Suppression	Planar Total	Dektor Total
Skylab III	2.20	2.30	2.11	1.91[*]	2.14	2.59[**]
Skylab IV	2.23	2.30	2.27	2.37	2.29	2.43

*   Significant - p<.001

**   Significant - p<.05

29

Thus, it can be seen that only one of the Planar sub-scores sub-
stantiated the hypothesis relative to this variable. The Dektor Total
score not only did not support the hypothesis but the obtained difference
was significant in the opposite direction.

Mission Day

The presumption here was that there might be systematic changes in
voice tracing scores as a function of time into the mission due to cumu-
lative fatigue or psychological stress. Of particular interest was the
apparent build-up in Mission IV of a series of misunderstandings between
crew members and ground personnel. These misunderstandings were resolved
on Day 46. Thus, a reduction in voice tracing scores following that day
would be considered indicative of a covariance of voice patterns with
what was probably a quite significant psychological change in the crew
members.

Table 5 presents the mean scores for each day of each mission on the
dependent variables (scores).

Table 6 summarizes the results for the early vs. the later segment
of each mission.

An examination of Table 5 indicates that Planar sub-scores and
Total score on day 47 of Skylab IV dropped dramatically below those ob-
tained for previous days. Inspection of Table 6 reveals that for Planar
scores all comparisons for Skylab IV and two of those for Skylab III show
statistically significant differences in scores in the later segments of
the missions with the later segment being lower in each case.

The fact that voice tracing scores were lower in the last part of
the mission would not support the hypothesis that there was a build-up of
stress throughout the mission. It is, of course, equally likely that,

30

TABLE 5

COMPARISON OF SCORES BY DAY OF MISSION

Mission Day	Blocking	Diagonal	Leading Edge	Amplitude Suppression	Planar Total	Dektor Total
SKYLAB III DAY						
18	2.49	2.38	1.64	2.28	2.20	
25	1.94	2.23	2.11	1.60	1.97	2.80
32	2.68	2.56	2.43	2.28	2.49	2.49
39	2.11	2.28	1.88	2.02	2.07	2.44
46	1.89	2.12	2.28	1.67	1.99	2.66
53	2.08	2.22	2.31	1.90	2.13	
SKYLAB IV DAY						
18	2.58	2.83	2.73	2.38	2.63	
25	2.23	2.30	2.55	2.85	2.49	
32	2.34	2.53	2.81	2.99	2.67	
39	2.24	2.07	1.79	1.83	1.98	
45	2.61	2.62	2.33	2.50	2.52	2.30
46	2.37	2.22	2.09	2.70	2.35	2.44
47	1.59	1.73	1.83	2.03	1.80	2.49
53	2.21	2.51	2.35	1.97	2.26	
60	2.00	2.16	2.04	2.18	2.10	
67	2.48	2.40	2.37	2.28	2.38	
74	1.84	1.96	2.11	2.42	2.08	2.53

with adaptation and practice, the latter portion of the mission was less stressful. Thus, it is not possible to state categorically on the basis of one comparison that the scores do or do not vary systematically with operational stress factors. The reversal in the results for Dektor Total Score is thought to reflect differences in scoring criteria. These will be discussed in a later section of the report.

31

TABLE 6

COMPARISON OF SCORES FOR EARLY AND
LATER SEGMENT OF EACH MISSION

Mission Segment	Blocking	Diagonal	Leading Edge	Amplitude Suppression	Planar Total	Dektor Total
SKYLAB III						
Days 18-32	2.37*	2.39*	2.06	2.05	2.22	2.65
Days 39-53	2.03	2.21	2.16	1.86	2.06	2.55
SKYLAB IV						
Days 18-46	2.40*	2.43*	2.38*	2.54*	2.44*	2.37*
Days 47-74	2.02	2.15	2.14	2.18	2.12	2.51

* Significant - $p < .05$

Time on Duty

It will be recalled that the Skylab crew members worked long (usually 16 hours) days under considerable pressure from a heavy schedule of experiments. Thus, it was felt that of all independent variables in the study, the one which could most unequivocally be said to represent stress was this. Previous research on demanding tasks performed over periods exceeding eight hours in duration have shown rather consistent cyclical within-day trends on a variety of physiological measures, e.g., Chiles, et al. 1968. These measures have included heart rate, skin resistance, skin temperature, and axillary temperature. In addition, Hale et al., 1971, reported significant changes in a variety of hormonal secretions as a function of shift length of air traffic controllers. Thus, the "time on-shift" of the Skylab crew members offered perhaps the most critical

independent variable in the present analysis. Table 7 presents the results for this variable. Data have been combined into four-hour blocks of time throughout the day since an hour-by-hour grouping yields a very small number of observations per cell.

TABLE 7

COMPARISON OF SCORES FOR TIME ON DUTY

Time on Duty	Blocking	Diagonal	Leading Edge	Amplitude Suppression	Planar Total	Dektor Total
1-4 Hours	2.21	2.27	2.11	2.09	2.17	2.48
5-8 Hours	2.07	2.21	2.22	2.20	2.18	2.50
9-12 Hours	2.30	2.31	2.22	2.30	2.28	2.43
13-16 Hours	2.34	2.50	2.48	2.51	2.46	2.68

An analysis of variance for each column indicates that in no case was there a statistically significant trend in voice tracing scores as a function of time on duty.

Since this variable was considered very important to the overall evaluation of the voice analysis technique, several additional analyses were performed, e.g. regression analyses of all dependent variables on Time on Duty and correlational analyses. In these analyses Time on Duty was structured in 16 intervals representing actual hours throughout the work day. Again no significant relationship was found between Time on Duty and any dependent variable.

The lack of positive relationships between voice tracing scores and the length of time crew members were on duty each day is considered the most important single indication that this technique is unsuitable in the

present application, since all available research information would lead to the presumption that conventional physiological measures would almost certainly vary systematically with this variable.

Task

The two tasks involved were ATM (Apollo Telescope Mount) and EREP (Earth Resources Experiment Package). EREP tasks were judged to be generally less demanding than ATM tasks, thus allowing an evaluation of the effects of task difficulty on tracing scores. Table 8 presents these results.

TABLE 8

COMPARISON OF SCORES AS A FUNCTION OF TASK

Task	Blocking	Diagonal	Leading Edge	Amplitude Suppression	Planar Total	Dektor Total
ATM	2.23	2.32	2.21	2.18	2.24	2.51
EREP	2.18	2.24	2.22	2.34	2.25	2.53

In no case was the difference in scores statistically significant. This finding is not considered of major consequence since the degree of stressfulness of the two tasks cannot be documented with confidence.

Activity

Here the comparison was between scores in a situation where the crew member was communicating while actually performing the experiment and one in which the task had been completed earlier and the crew member was

making an oral report of procedures and data. The supposition was that
performing was more stressful than reporting. Table 9 presents the data
relevant to this comparison.

TABLE 9

COMPARISON OF SCORES AS A FUNCTION OF ACTIVITY

Activity	Blocking	Diagonal	Leading Edge	Amplitude Suppression	Planar Total	Dektor Total
Performing	2.21	2.31	2.17	2.10*	2.20	2.52
Reporting	2.22	2.29	2.26	2.35*	2.28	2.52

* Difference between means significant - p<.05

It can be seen that, for one sub-score, Amplitude Suppression, the
difference between performance and reporting activities was statistically
significant. The difference was, however, in the unexpected direction.
Thus, the hypothesis was not supported. As with the previous variable,
"Task," the authors are not prepared to make a strong case that there is
an important difference in stressfulness between performing and reporting
as defined in this study.

Crew Position

The Commander, Science Pilot, and Pilot in each mission had differing
general duties and responsibilities. For the tasks studied here (ATM and
EREP) they had essentially identical responsibilties, however. The com-
parison here is made to determine whether or not voice tracing scores vary
systematically as a function of position and general responsibility. Table
10 presents these results.

TABLE 10

COMPARISON OF SCORES AS A FUNCTION OF CREW POSITION

Crew Position	Blocking	Diagonal	Leading Edge	Amplitude Suppression	Planar Total	Dektor Total
Commander	2.13	2.34	2.19	2.20	2.21	2.50*
Science Pilot	2.19	2.20	2.19	2.15	2.18	2.39*
Pilot	2.33	2.36	2.27	2.34	2.32	2.66*

* ANOVA Significant - p<.01

Again one measure, the Dektor Total score, relates significantly to the independent variable. If several measures had shown this relationship, the usefulness of the technique as an aid in the distribution of responsibilities among crew members and other similar considerations would have been indicated. With only one such indicator, the relationship between tracing scores and the position of the crew member is so weak as to make decisions based on those scores indefensible. Of course, the small number of incumbents for each position (two) would have made any findings suggestive at best.

Speaker

The primary importance of this portion of the analysis was to make certain that, if positive findings on the other independent variables emerged, they were not purely a function of individual differences among speakers. Table 11 presents the findings.

The only measure on which scores varied significantly as a function of the speaker was the Dektor Total score. (This relationship probably accounts for most of the variance in the previous comparison -- Crew

TABLE 11

COMPARISON OF SCORES AS A FUNCTION OF SPEAKER

Speaker	Blocking	Diagonal	Leading Edge	Amplitude Suppression	Planar Total	Dektor* Total
1	2.00	2.04	2.06	1.88	2.00	2.38
2	2.21	2.16	1.99	2.00	2.09	2.48
3	2.38	2.69	2.28	1.99	2.33	2.93
4	2.20	2.50	2.26	2.37	2.33	2.62
5	2.18	2.23	2.29	2.23	2.23	2.31
6	2.30	2.18	2.27	2.53	2.32	2.39

* ANOVA Significant- $p < .001$

Member). Thus, scored by the Planar scoring procedures described earlier in this report, individual differences in vocalization did not have a material effect.

Secondary Dependent Variables

It will be recalled that two variables which might have had an effect on voice tracings were included in the study. These were:

- Voice Quality - The technical quality of the tape
recording as judged subjectively,
- Recording Mechanism - Air-to-ground vs. tape-dump.

Table 12 presents the results for Voice Quality, Table 13 the results for Recording Mechanism.

The implication of finding three scores which vary significantly

37

TABLE 12

COMPARISON OF SCORES AS A FUNCTION OF VOICE QUALITY

Voice Quality	Blocking*	Diagonal*	Leading Edge	Amplitude Suppression	Planar Total	Dektor** Total
Good	2.33	2.54	2.34	2.32	2.38	2.84
Fair	2.22	2.30	2.28	2.16	2.24	2.50
Poor	2.11	2.10	1.99	2.28	2.12	2.32

   *   ANOVA significant – p<.05
  ** ANOVA significant – p<.001

TABLE 13

COMPARISON OF SCORES AS A FUNCTION OF RECORDING MECHANISM

Recording Mechanism	Blocking	Diagonal	Leading Edge	Amplitude Suppression	Planar Total	Dektor Total
Dump-tape	2.19	2.27	2.21	2.21	2.22	2.53
Air/Ground	2.35	2.43	2.23	2.32	2.33	2.40

with judged voice quality is of considerable importance. It can be seen that scores were lower in each case for the tape recordings of poorer quality. Apparently this poor quality, which results in many irrelevant pen movements as the chart is being produced, tends to obscure patterns of the type scored here. The implication is, of course, that in future applications one should obtain recordings of the highest possible quality if they are to be used for voice analysis.

The findings for the type of recording mechanism used are presented in Table 13. None of the differences in this comparison is

significant. While there are minor differences between recordings made on various tape recorders and between generations when copies of a tape are made, these generational differences do not appear to affect systematically the patterns scored in this study.

Correlational Analyses

It is useful to examine the interrelationships of the dependent variables. This is especially true since the scoring methods used in this study were extremely tedious and time-consuming. If it can be shown that a smaller number of scores or a less rigorous set of scoring procedures produce essentially the same results as the detailed procedures used here, significant savings in time and effort can be made in future studies. Table 14 presents a correlation matrix for the primary dependent variables on which the preceding analyses were based.

TABLE 14

CORRELATION COEFFICIENTS – ALL PRIMARY DEPENDENT VARIABLES

Scores	Blocking	Diagonal	L. Edge	Amp. Supp.	Planar Total	Dektor Total
Blocking	---	.80	.41	.40	.85	.41
Diagonal		---	.51	.28	.85	.46
Leading Edge			---	.28	.71	.36
Amplitude Suppression				---	.65	-.03
Planar Total					---	.39
Dektor Total						---

As would be expected, the correlation coefficients for the sub-scores and the Planar Total are quite high. This is another indication of the intra-judge reliability of the scoring since these sub-scores, while contributing to the Total Score, were assigned independently of one another.

The low correlation between the Planar Total score and the Dektor Total score probably reflects two primary differences in scoring technique. The Dektor Total score was assigned in each case by the Chief Instructor of Dektor Counterintelligence and Security, Inc. His very considerable experience in this area is almost entirely in analyzing voice tracings obtained in a highly structured interview situation where he knew the speaker, the situation, and the content. Thus, his typical requirement is the examination of variations in patterning within a given chart rather than blind analyses of individual phrases. In addition, the Planar scoring placed an equal weight on Amplitude Suppression with the other scores. The Dektor scoring did not consider this variable in such an important way; thus, the lack of correlation between the Amplitude Suppression score and the Dektor Total score.

The other correlational analysis of interest examined the relationship between estimates of patterning obtained after a very brief examination of the entire (20 episode) chart and scores assigned on the basis of meticulous scoring of each of the 20 episodes against rigid rules.

It can be seen that the estimated scores relate quite closely to those obtained from meticulous and time-consuming scoring. Thus, for applications where very large numbers of voice charts must be considered, an estimate of each of the scores can safely be substituted for detailed scoring without serious loss of reliability. (See Table 15.)

TABLE 15

CORRELATION COEFFICIENTS AMONG ESTIMATES AND SCORES

Variables	Correlation Coefficients
Blocking Estimate and Blocking Score	$r = .87$
Diagonal Estimate and Diagonal Score	$r = .89$
Leading Edge Estimate and Leading Edge Score	$r = .85$
Amplitude Suppression Estimate and Amplitude Suppression Score	$r = .81$
Overall Estimate and Planar Total Score	$r = .78$

CHAPTER IV

CONCLUSIONS

The Dektor Psychological Stress Evaluator (PSE) makes a graphic record
of signals produced by the human voice. This record is capable of
reliable classification into categories which are themselves relatively
independent of one another. The characteristics of the graphic record are
to some extent influenced by the quality of the tape recording from which
it is made. Extremely detailed scoring does not appear to be justified
in situations such as the one evaluated here. Careful, but brief review
of the graphic records enables one to determine the patterns of interest
with high accuracy as compared with detailed scoring.

The technique, as used in the manner described above, does not appear
to measure variables which relate to an operationally useful extent to
the relatively mild gradations of stress involved in the performance of
the Skylab tasks included in this study.

It must be remembered that the device is not commonly used in an
application such as this. No inferences should be drawn as to its use-
fulness in highly structured interview situations where within-subject
and within-interview change in voice patterns is the important consideration.

Based on the findings, it is not recommended that this voice analysis
technique be used, in the manner described in this report, as an indicator
of the stressfulness of tasks or activities in manned space missions. It
is recommended that continuing efforts be applied to the development of
other techniques for the processing of voice communications, e.g., formant
analysis.

APPENDIX A

DESCRIPTION OF PSE VOICE ANALYZER

The Dektor PSE Voice Analyzer is manufactured by Dektor Counterintelligence and Security, Inc., 5508 Port Royal Road, Springfield, Virginia. It was developed in response to the requirement for an advanced interrogation capability which does not involve the use of attached sensors and which can be used in a relatively uncontrolled environment. The PSE can be used without inducing artificial stress in the subject by the testing environment, and it will allow testing to be accomplished over a remote communication link or from any voice recording medium.

The voice analyzer functions by detecting and processing selected subaudible voice frequencies which change in a predictable manner as a result of psychological stress. As such, it provides a means of accurately determining and recording degrees of psychological stress in the speaker at the time of utterance.

As delivered by the manufacturer, the voice analyzer consists of three major components; the input device (which is a standard off-the-shelf magnetic tape recorder), the analyzer itself, and the output device (a standard off-the-shelf strip chart recorder). Of interest in this discussion is the voice analyzer device, a simplified block diagram of which is shown in Figure A-1. The principal components of the voice analyzer are as follows:

Low Pass Filter

This is an inductor, whose value has been chosen to attenuate frequencies above 30 Hz. It is connected in series between the recorder output and the rectifier (or bypassed completely) by means of operator control switches which also select low pass filtering at the output of the rectifier.

Rectifier

The rectifier produces a DC output level proportional to the AC energy present in the input signal from the magnetic tape recorder.

Magnetic Tape Recorder Output → Low Pass Filter → Rectifier → Low Pass Filter/ DC Level Converter → Operational Amplifier

Gain and Zero Controls

Chart Recorder Driver Amplifier

From Chart Recorder

To Chart Recorder
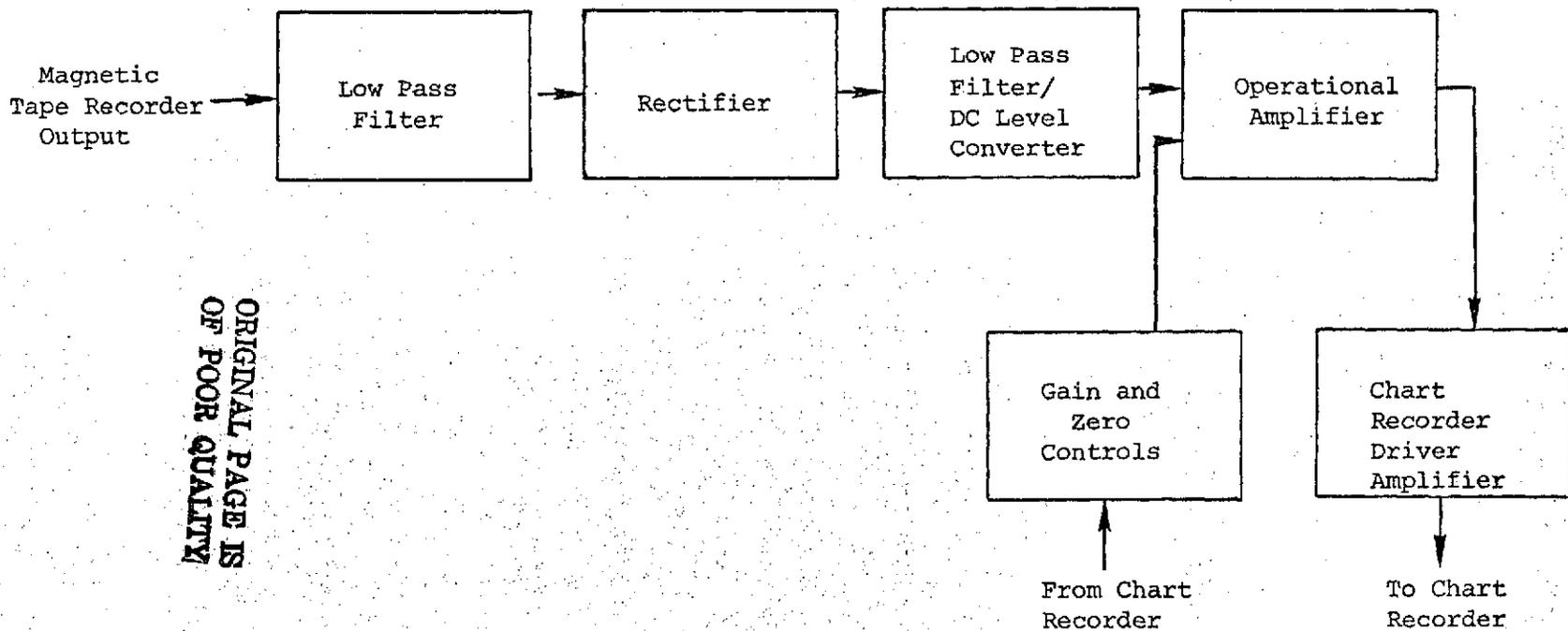
Figure A-1. Voice Analyzer Block Diagram

Low Pass Filter/DC Level Converter

The amount of low frequency energy passed by this component is dependent upon the setting of the operator control switches which connect the rectifier output to ground through various capacitors. The capacitor selected by the control switches also determines the point at which a DC level begins to become felt at higher frequencies.

Operational Amplifier

The operational amplifier, in conjunction with the gain and zero controls, amplifies and controls the rectifier output.

Gain and Zero Controls

These are potentiometers which provide feedback to the operational amplifier from the chart recorder to control the zero position and the extent of travel of the chart recorder pen.

Chart Recorder Driver Amplifier

The operational amplifier output is fed to this power amplifier to provide the signal to drive the pen of the chart recorder.

The signal processing characteristics of the voice analyzer are summarized in Figures A-2, and A-3. Figure A-2 shows the manner in which the AC output of the voice analyzer is dependent upon input frequency. The 3 db point of the frequency response is between 2 and 12 Hz, depending upon the setting of the operator control switches. Figure A-3 shows how the DC level of the voice analyzer is dependent upon the frequency of the input signal. Frequencies above approximately 50 Hz (again, dependent upon the control settings) begin to appear on the output of the voice analyzer as a DC level.

In actual operation, the response curves combine to produce an output signal in which low frequency components of the voice input are relatively unchanged, and higher frequency components appear as a DC level. Analysis of the output signal as it appears on the chart recorder tracing deals with several characteristics of the trace. Figure A-4 shows these characteristics.

3db

AC
OUTPUT

2-12 Hz
INPUT FREQUENCY
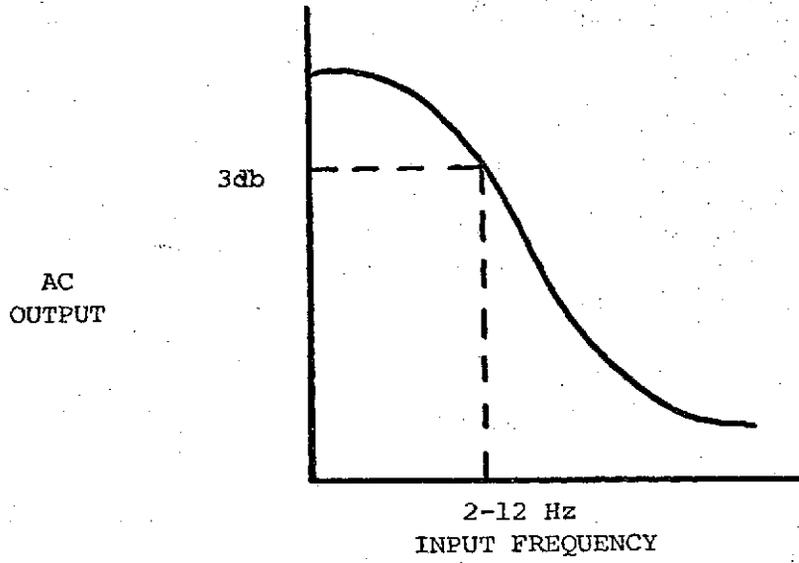
Figure A-2. Voice Analyzer Frequency Response

DC
OUTPUT

≈50 Hz
INPUT FREQUENCY
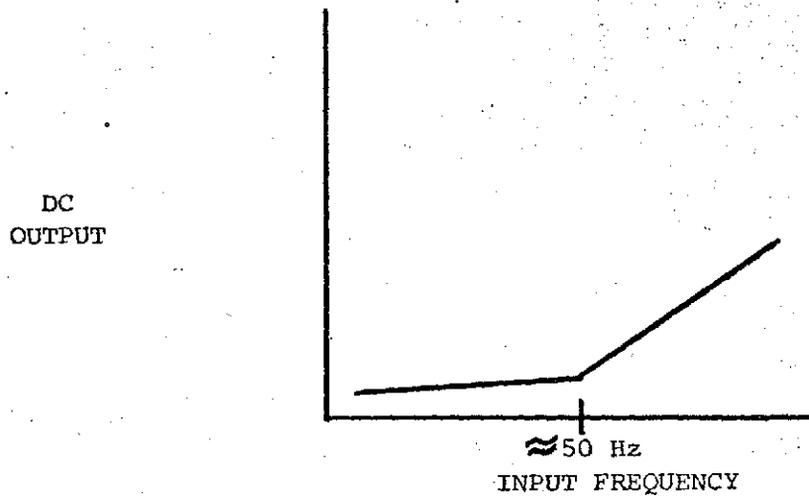
Figure A-3. Voice Analyzer Input Frequency vs. DC Output Level

Analysis of the output is concerned with the quantities A, B, and C as shown in Figure A-4. A is a measure of the leading edge of the signal. It is examined for steepness and smoothness. B, or 1/B, is a measure of the frequency of the signal. It is examined for discrete changes which can occur approximately midway between the leading and trailing edge of the signal. It is examined for discrete changes which can occur approximately midway between the leading and trailing edge of the signal. C is a measure of the DC level of the signal. It is examined for its rate and manner of change over the duration of the signal. The amplitude of the signal and the trailing edge of the signal are of no concern in the analysis.
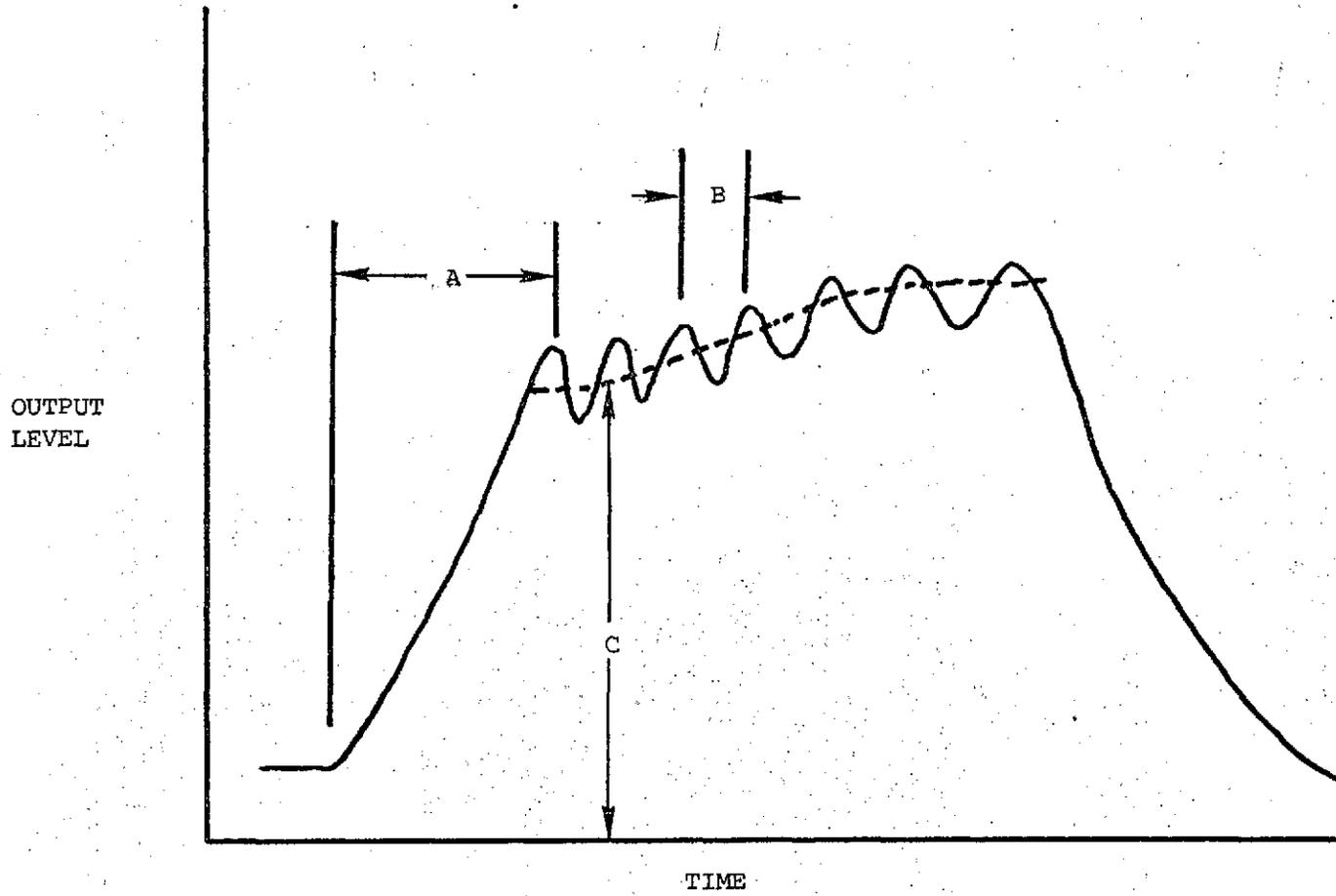
Figure A-4. Output Signal Characteristics

APPENDIX B

BIBLIOGRAPHY

Chiles, W.D., Alluisi, E.A., Adams, O.S. Work schedules and performance
during confinement. Human Factors, 1968, 10(2) 143-196.

Friedhoff, A.J., Alpert, M. and Kurtzberg, R.L. An electro-acoustic
analysis of the effects of stress on voice. Journal of Neuropsychiatry
1964 5(5) 266-272.

Grether, C.B. Voice output as a psychophysiological index of fatigue,
unpublished paper, North Carolina State University, Raleigh, North
Carolina (USAF Office of Scientific Research, Contract #F44620-60-C-
0033) 1971.

Hale, H.B., Williams, E.W., Smith, B.N., and Melton, C.E. Jr. Excretion
patterns of air traffic controllers. Aerospace Medicine, 1971 4(2)
127-138.

Hecker, H.M.L., Stevens, K.N., VonBismarck, G. and Williams, C.E. Mani-
festations of task-induced stress in the acoustic speech signal.
Journal of the Acoustical Society of America, 1968 44(4) 993-1001.

Huttar, G.L. Relations between prosodic variables and emotions in normal
American English utterances. Journal of Speech and Hearing Research,
1968 11(3) 481-487.

Lindsley, D.B. Emotion. In S.S. Stevens (Ed.) Handbook of Experimental
Psychology. New York: John Wiley (1951).

Popov, V.A., Simonov, P.V., Frolov, M.V. and Khachatur'yants, L.S. Fre-
quency spectrum of speech as an indicator of the degree and nature
of emotional stress. Zhurnal Vysshey Nervnoy Deyatel'nosti, 1971
21.Moscow (in Russian) translation Joint Publications Research
Service, U.S. Department of Commerce, March 1971.

Simonov, P.V. and Frolov, M.V. Utilization of human voice for estimation
of man's emotional stress and state of attention. Zhurnal Vysshey
Nervnoy Deyatel'nosti, 1971, 21. Moscow (in Russian) translation
Joint Publications Research Service, U.S. Department of Commerce,
March 1971.

Sternbach, R.A. Principles of psychophysiology. New York: Academic
Press (1966).

Waskow, I. The effects of drugs on speech: a review. Psychopharmacological
Bulletin, 1966 3(4) 1-20.

Williams, C.F., and Stevens, K.N. On determining the emotional state of
     pilots during flight: an exploratory study. <u>Aerospace Medicine</u>,
     1969 40(12) 1369-1372.