NASA Contractor Report 4258

Voice Measures of Workload in the Advanced Flight Deck: Additional Studies

Sid J. Schneider and Murray Alpert

CONTRACT NAS1-18278
NOVEMBER 1989

NASA Contractor Report 4258

Voice Measures of Workload in the Advanced Flight Deck: Additional Studies

Sid J. Schneider and Murray Alpert
Behavioral Health Systems, Inc.
Ossining, New York

Prepared for
Langley Research Center
under Contract NAS1-18278

National Aeronautics and
Space Administration
Office of Management
Scientific and Technical
Information Division

1989

Abstract

These studies applied the results from the first report (Schneider, Alpert, & O'Donnell, 1989) to voice samples obtained from individual operators. In study 1, one person performed a task, similar to one described in the first report, in which voice samples were recorded in high, medium, and low workload conditions. The results suggested that for this single individual, mean amplitude, frequency, peak (syllable) duration, and stress (emphasis) all tended to increase as workload increased. In study 2, NASA test pilots performed the same task. They also used a flight simulator under high and low workload conditions while their voices were recorded. The results from the simulator suggested that for two of the pilots, high workload brought about greater amplitude, peak duration, and stress. In both the laboratory and simulator tasks, high workload tended to be associated with more statistically significant drop-offs in the acoustical measures than were lower workload levels. The acoustic measures displayed a great deal of variability, both among subjects, and within the samples from individual subjects. These results are discussed as they pertain to the use of voice measures to assess the operator demands imposed by new technology.

This study was intended to extend the work described in the first report (Schneider, Alpert, & O'Donnell, 1989), by evaluating whether acoustical analysis of the voice can measure the workload experienced by individual test pilots. The study described in the first report used a group of non-pilots, who performed a laboratory task that was not directly related to piloting an aircraft. Voice samples were recorded while the workload level was systematically manipulated.

The results suggested that the mean amplitude and frequency of the subjects' voices were greater in the high workload condition than they were in the low workload condition. These differences were not statistically significant. This result was similar to results reported by Shipp, Brenner and Doherty (1986). Further analyses revealed that in both workload conditions, the amplitude and frequency of the voice diminished over time, perhaps a reflection of the subjects' fatigue as the tasks went on. The drop-off in amplitude and frequency was significantly greater in the high workload condition. This result may suggest that energy is lost from the voice most quickly when the task demands upon the speaker are greatest.

The results further suggested that there was a great deal of variability in the acoustical parameters of the voice, not only among the different subjects, but also within the samples obtained from each individual subject. This intra-subject variability called into question the utility of voice parameters as a measure of the workload experienced by one single individual. When voice recordings are collected from any single individual, the effects of workload may be masked by fluctuations in the voice unrelated to workload. The voice is under voluntary control; individuals might even "correct" for the effects of workload by deliberately raising the volume and frequency of their voices as a task wears on. It may be necessary to obtain voice recordings from a relatively large subject sample for the effects of workload to become apparent.

The present studies were intended to reveal the feasibility of assessing the workload experienced by individual pilots through measurements of the acoustical properties of their voices. The first study was intended to demonstrate whether data from a single subject, collected under controlled conditions, could show the effects of workload. In the second study, three NASA test pilots were recorded while using a flight simulator to "land" an aircraft four times. The demands of each landing were manipulated by changing the crosswinds and turbulence. Subjective ratings were obtained from the pilots to confirm that these weather changes had the intended effect.

In addition, the three pilots also performed a laboratory task very similar to the one described in the first report (Schneider, Alpert & O'Donnell, 1989). The laboratory task was intended to reveal how workload affected the voice of the individual pilots under controlled conditions. For example, the amplitude of one pilot's voice might fall over time more in a high workload condition than in a low workload condition; for another pilot, it may be frequency that is most affected by workload. These profiles of the individual effects of workload could then be applied to the voice samples obtained in the simulator. The effects of workload under the controlled laboratory task might be replicated in the simulator task.

Study 1

Method

The purpose of this study was to determine whether a modified version of the laboratory task described in the first report could be used to assess the workload of a single operator. The earlier work examined only mean values obtained from a group of subjects. In the present study, there was only one subject.

The subject performed a laboratory task very similar to the one described in the first report. In order to eliminate learning effects that had been observed in the earlier study, the subject performed the task six times, first once on a Friday, and then once daily on Monday through Friday of the following week. The voice samples collected on Thursday, the next-to-last day of the study, were analyzed. In this way, the subject had a great deal of experience with the task and was no longer learning it when his samples were recorded. The next-to-last day was used rather than the last day to avoid any letdown that might occur on the last day of the study.

The details of the task are described in the earlier report (Schneider, Alpert, & O'Donnell, 1989), but will be briefly summarized here. Voice samples were obtained by requiring the subject to speak whenever one of two triangles on a computer monitor began to rotate. At random intervals ranging from 20 to 25 seconds, one of the triangles would rotate and the subject was required to say in his normal voice, "Triangle please stop turning now." The triangle actually did stop spinning when the subject stopped speaking. Subjects wore headphones which presented white noise at 60 dB (0.0002 microbar reference) to simulate flight deck sounds and to mask noise outside the laboratory.

There was a secondary task whose purpose was to vary the overall workload. The secondary task was a version of the Continuous Performance Test (Rosvold, Mirsky, Sarason, et al, 1956) in which numerals were presented, one after the other, in the center of the computer screen. The numerals 1 through 6 were used. The subject was required to press a button, which he held in his hand, as quickly as possible whenever two successive numbers added to 7. The computer software arranged the numbers in a random sequence so that 30 percent of the numbers required button pushes. The software also recorded the number of omission, commission, late, and double strike errors. From those figures, the software continually calculated the error rate, and adjusted the speed at which the numerals were presented to hold the error rate as constant as possible. In this study, the error rate was held to .30 in the low workload condition, .50 in the moderate workload condition, and .70 in the high workload condition.

In the earlier study, there were only two workload levels, in which the error rates were respectively .20 and .60. These levels were modified because two subjects could not improve their performance to the .20 level regardless of how slowly the numerals were presented. Also, by adding a moderate workload level, it would be possible to more clearly observe trends in the effects of workload. The updated software used in the this study used subroutines that provided more accurate timing and clearer graphics.

There were 14 runs, each eliciting 12 voice samples. There was a one-minute rest period after each run. The runs were presented in the order BLMHHMLLMHHMLB, in which B stands for "baseline" (no continuous performance task at all), and L, M, and H respectively stand for the low, moderate, and high workload conditions.

The methods for performing the acoustical analyses were as described in the first report. The hardware used in for the analyses was updated. The Northstar computer was replaced with an IBM PC/AT compatible.

Results and Discussion

Table 1 shows the mean amplitude, frequency, peak (syllable) duration, and stress (emphasis, a function of the other three variables) of the four workload conditions. The results suggest that there was a general increase in all four acoustical parameters as workload increased. The exception to this trend was the low workload condition, for which all acoustical parameters except frequency were somewhat higher than those for the moderate workload condition.

Further inspection of the data revealed that the reason for this break in the trend was that in the third low workload run, the subject's speech had uncharacteristically elevated amplitude, frequency, and duration (see Figs. 1 & 2). It appears that as the task went on, the subject may have become fatigued. He may have deliberately injected new energy into his voice in the eighth run, which was the third low workload run. Because of that single run, the acoustical measures for the low workload condition were elevated. After that run, the energy of the subject's voice declined. The amplitude and frequency reached low levels by the last three runs (see Figs. 1 & 2). It appears that, after doing the task six days in a row, the subject knew when to anticipate the end of the task, and allowed the energy in his voice to wane as the end approached.

These results suggest that increased workload brought about increased amplitude, frequency, duration and stress, at least in the voice of this one subject. The repeated administrations of the task apparently succeeded in removing any learning effect; the data concerning the subject's performance on the continuous performance task do not point to a general improvement in performance across the runs. The removal of the learning effect may have made the workload effect more conspicuous. However, the subject may have overridden the effects of workload by increasing amplitude, frequency, and stress, at least during the eighth run.

The temporal effects of workload that had been apparent in the earlier study were not apparent in the data for this subject. Amplitude, frequency, duration, and stress all tended to fall across the twelve voice samples collected in each workload condition. However, the drop-offs were no greater in the high workload condition than any other condition. For this particular subject, the differences among the workload conditions were apparent in the mean values of the acoustical measures, not in how the measures diminished over time.

In order to more clearly establish how the acoustical measures changed over the course of each run, Pearson product moment correlations were calculated between the acoustical measures and the serial position of the twelve utterances. Several of these correlations were statistically significant for the amplitude measure (and stress as well, since amplitude is a factor in the stress measure). The correlation calculated for the first baseline run, -.83 ($p < .001$) suggests a large drop-off. This was the first trial of the day. The subject may have began the task speaking unusually loudly, and his voice became less loud as the run went on. The correlation for the eighth run, the low workload run whose mean was unusually high, was -.70 ($p < .01$), suggesting it too had a large drop-off. The finding supports the idea that the subject may have temporarily injected new energy into his voice at this point. The correlations for runs 12 (medium workload) and 13 (low workload) were respectively -.68 ($p < .02$) and -.55 ($p < .10$). These runs were among the last of the day, when the subject's voice was reaching low mean amplitude levels. The finding again is consistent with the idea that these drop-offs reflect fatigue.

The earlier study revealed a great deal of variability among the subjects in the temporal effect of workload. Because of such individual differences, it may be difficult to use acoustical voice analysis based on group results to assess the workload experienced by a single operator. Perhaps, by determining first how the individual's voice is affected by workload in a controlled, laboratory task, it can be possible to predict how that individual's voice will be affected by workload in an actual work environment. The following study was to examine that hypothesis using three NASA test pilots, who both performed the laboratory task, and "landed" an aircraft under two workload conditions in a flight simulator.

Study 2

Method

Subjects. There were three subjects, each a male NASA test pilot who was familiar with the flight simulator used in this study.

Laboratory task. The subjects were run individually in the same laboratory task that was used in study 1 above. It was not possible to run the test pilots six days in a row. Therefore, the data from study 1 was inspected again to determine when the effects of learning diminished. It appeared that learning effects had greatly diminished after the first day's runs in study 1; that is, the subject's performance on the continuous performance test did not improve across runs after the first day. Therefore, the test pilots were run on two consecutive days, and the data from the first day were discarded.

The laboratory task was shortened to only ten runs, in the order BLMHHMLLMH. During study 1, there had been 14 runs: those ten, followed by HMLB. The final four runs were now omitted to reduce the effects of fatigue caused simply by the length of the task, not workload itself. As noted above, such fatigue may have influenced the results from study 1. In the present study, the initial, baseline run was intended to serve as practice, to reduce any effect for the novelty of the task. In study 1, data from the first run of the day had suggested an unusually steep reduction over time, probably unrelated to workload. The last, high workload run was lost from the data for subject 3 due to a technical problem.

Simulator task. On a separate day, the subjects "landed" a Boeing 737 aircraft four times at Langley Air Force Base in a NASA flight simulator, in velocity control wheel steering mode, using an instrument landing system approach. In the low workload condition, there were no crosswinds or turbulence. In the high workload condition, there were moderate crosswinds and turbulence (10 knots each), about as severe as found in a summer storm. Subject 2 reported that he noticed no additional difficulty from the added crosswinds and turbulence. Therefore, he landed the aircraft using a manual throttle in the high workload condition.

The other subjects used automatic throttle in both conditions. The workload conditions were presented in the order LHHL.

During each run, there was a buzzer near the subject in the simulator. The buzzer sounded every 90 seconds; since each run lasted about 15 minutes, this was about 10 times per run. The subject was required to report his subjective rating of the difficulty level of the procedure that he was performing at the moment, on a scale of 1 to 10 where 1 stood for trivially simple and 10 for extremely challenging.

A continuous recording was made of the subject's speech throughout each run. When these recordings were acoustically analyzed, only the subject's communications with the air traffic controller were included. All other verbalizations were edited out.

Results

Laboratory task. Tables 2 through 4 show the mean values for the acoustical measures obtained from each subject in the low, moderate, and high workload conditions. None of the parameters, for any of the three subjects, increased systematically as workload increased. For each subject, the values for the high workload condition were higher than those of the low workload condition for only one or two of the four parameters--about as many as could be attributed to chance. It appears that workload had no systematic effect on the mean values of the acoustical measures in the laboratory task.

Further analyses examined the drop-offs over time in amplitude, frequency, duration, and stress, to determine whether increased workload brought about greater reductions over time in any of these measures. The mean values for each measure were calculated for the first three and the last three voice samples in the low, moderate, and high workload conditions. Examination of the data revealed that drop-offs did occur. There were three workload conditions, and four acoustical measures for each subject. Therefore, there were twelve measures of change over time for each subject. For subjects 1 and 2, 9 of the 12 changes between the first and last three voice samples were drop-offs; for subject 3, 10 of the 12 changes were drop-offs.

The first hypothesis to be examined was that the greatest difference between the first and last samples would occur in the high workload condition, followed by the moderate and then the low workload conditions. Examination of the data failed to support the hypothesis. The magnitude of the drop-offs between the first three and the last three voice samples was not related to workload level for any subject.

The next hypothesis to be examined was the one suggested by the results in the earlier report. There were three runs in each of the three workload conditions. The results from the earlier study suggested that the drop-offs over time would increase fastest from the first to the third run in the high workload condition. In other words, the first run in the high workload condition might show a small drop-off in the acoustical measures; the second run might show a larger drop-off and the third run might show a yet larger drop-off. This trend would be weaker in the moderate workload condition, and weakest in the low workload condition. Again, the data did not support the hypothesis. In fact, the drop-offs in the acoustical measures did not increase across runs, even the high workload condition, for any subject.

The data for the baseline condition, which was intended as a practice run, was also inspected. The levels of the acoustical parameters, and the extent of their drop-offs, were not systematically lower than the corresponding values from the workload conditions.

In summary, inspection of the data from the three subjects failed to suggest that workload had any systematic effect upon any acoustical measure. However, the measures of drop-offs in the acoustical measures used only the first and last three utterances in the runs. These measures could give some indication of the magnitude of a drop-off, but could not quantify the relationship between the acoustical measure and the passage of time. In order to more clearly establish the degree of change in the acoustical measures over time in each workload condition, Pearson product moment correlations were calculated, as they were in study 1. For each run, the correlation was calculated between the acoustical measure and the serial position of the twelve utterances. The results for each acoustical measure, for each subject, in each run are shown in Tables 5 through 8.

The tables show that in general, there were drop-offs in all the acoustical measures in all workload conditions; most of the correlation coefficients (80 of the 116) were negative. Six of the correlation coefficients calculated for the amplitude data were significant at the .02 level; four of these were in the high workload condition. Of these four significant correlation coefficients, there were two for subject 2, and one for both subjects 1 and 3.

Only one correlation coefficient for the frequency and peak duration measures was significant at the .05 level among all the subjects. However, two correlation coefficients, one each for subjects 2 and 3, were significant at the .02 level for the stress measure. Both were in the high workload condition.

Simulator Task. Table 9 shows each subject's mean subjective ratings for the difficulty level (rated from 1 to 10) that he experienced throughout each run. Each subject provided between nine and eleven reports of the difficulty level in each run. The table shows the mean difficulty level of the first five and last five reports in the two workload conditions. It also shows the overall mean for each workload condition.

The table shows that the subjective difficulty level was greater during the last five reports as compared with the first five. During the last five reports, the pilots were into the descent and touchdown on the runway. During the first five reports, the pilots were still approaching the airport, a procedure that all three pilots found less demanding.

The table also shows that the addition of turbulence and crosswinds, and, for subject 2, manual throttle, had their intended effect for all subjects. The difficulty ratings were greater in the high workload condition than in the low workload condition. The magnitude of this difference in the workload conditions was greatest for subject 2, least for subject 1.

Table 10 shows the mean acoustical values from the voice samples recorded in the simulator for each subject in each workload condition. The values were obtained from the first six and the last six voice samples in each run. Thus, twelve values were obtained in each of the two runs in each workload condition for each subject. To obtain each value shown in the table, the 24 values obtained for each workload condition for each subject were averaged together. This procedure assured that an equal number of voice samples for each subject went into the analysis, both from the early part of the simulation, and from the more difficult late part. The subjects differed greatly in how many voice samples they provided while working in the simulator; averaging the values was intended to compensate for those differences.

The asterisks on Table 10 show where the high workload condition produced higher values than the low workload condition produced. High workload was not associated with increased frequency for any subject. However, high workload did bring about increased amplitude for all subjects, and increased peak duration and stress (emphasis) for subjects 1 and 3.

Further analysis examined the drop-offs in the acoustical measures over time. For each subject, the mean for each acoustical measure was calculated early in the run, later in the run, and at the end of the run. These calculations were done by recording the values for each acoustical measure for the first three voice samples in each run, for voice samples numbers 10, 11, and 12, and for the final three voice samples. Means were computed for each subject in each workload condition and are shown on Tables 11, 12, and 13. The tables also show the magnitude of the drop-off in each acoustical measure between the early and middle part of the simulation, between the middle and late part, and between the early and late part.

Table 11 (data for subject 1) shows that the values for all the acoustical measures fell between the early and middle parts of the simulation. The drop-off in amplitude was greater in the high workload condition than in the low workload condition. Between the middle and late parts of the simulation, there were small rises in the values for every acoustical measure in the low workload condition; in the high workload condition, the values for every acoustical measure fell. Thus, late in the simulated landing, drop-offs in every acoustical measure were observed only in the high workload condition. Across the entire length of the simulation, Table 11 shows that the drop-offs in all the acoustical measures except frequency were greater in the high workload condition than in the low workload condition.

Table 12 (data for subject 2) shows that there were drop-offs in every acoustical measure except peak duration, in both workload conditions, between the early and middle parts of the simulation. These drop-offs were greater in the high workload condition than they were in the low workload condition. The rise in peak duration was smaller in the high workload condition. This trend for greater drop-offs in the high workload condition was apparent only between the early and middle parts of the simulation; it was not apparent between the middle and late parts.

Table 13 (data for subject 3) also suggests greater drop-offs for every acoustical measure between the early and middle parts of the simulation. In fact, the only consistent drop-offs in subject 3's data are in the acoustical measures recorded in the high workload condition, between the early and middle parts of the simulation. All other changes in the subject's vocalizations were increases in the acoustical measures.

In summary, there was evidence in the data for each of the subjects for greater drop-offs in the acoustical measures in the high workload condition. However, this trend was not consistent. It occurred only early in the simulation for subjects 2 and 3, only late in the simulation for subject 1. There was even some evidence for greater drop-offs in the low workload condition for subject 1 early in the simulation, and for subject 2 late in the simulation.

In light of these inconsistent results, correlation coefficients were computed, similar to those obtained for the data collected in the laboratory task. The Pearson product moment correlation between the serial position of the utterance, and the acoustical measure was calculated for every run. The first six and the last six utterances in each run were used. The results are shown in Tables 14 through 17.

Most of the correlation coefficients (33 of 48) on the tables are negative, suggesting that the acoustical measures generally decreased over the course of the runs. For the amplitude measure, two of the three coefficients significant at the .10 level were in the high workload condition (one of two at the .05 level). For the stress measure, three of the four coefficients significant at the .10 level for stress were in the high workload condition (one at the .02 level). Two coefficients for frequency were significant at the .01 level in the high workload condition. No correlation coefficient for peak duration was statistically significant.

Discussion

There are several stages to the process by which workload could affect the acoustical properties of the voice. First, workload must affect the mental state of the speaker. The mental state must then affect the physical state, such as by changing muscle tension. These muscular changes then must affect speech production, for example by tightening or relaxing the vocal cords or altering the force with which the diaphragm contracts (Cannings et al, 1979). Finally, these changes in speech production must be reflected in the acoustical measures obtained through computer analysis of the voice.

There are many factors unrelated to workload which can complicate this process. Mental and physical states are affected by a range of factors which might obscure the effects of workload. The musculature involved in speech production is under voluntary control; any effect of workload can be overridden by the operator's speaking habits. Moreover, there is a great difference among speakers in the extent to which frequency, amplitude, and other measures can vary. Some voices have a wide range of frequencies and amplitudes, while others have a limited range (Cannings et al, 1979). Thus, acoustical measures of the voice are likely to reflect many processes in addition to workload at any time. In the present series of studies, there was a great deal of variability in the effects of workload on the acoustical properties of the subjects' voices. There was variability both among utterances from different subjects and among utterances from a single subject. This variability may reflect the many factors besides workload that affect the voice.

Despite this variability, the results generally suggest that increased workload brought about increased energy in the voice. In the earlier report, increased workload was associated with increased frequency, amplitude, and peak duration in 14 non-pilots, although the increases were not statistically significant. For the subject in study 1, and for the pilots in the simulator study, high levels of workload were associated with higher amplitude, stress, and peak duration. For the subject in study 1, frequency levels also went up as workload increased. However, the test pilots in study 2 never showed this effect for frequency. Also, in the laboratory task, the mean values of the acoustical measures of the pilots' voices were not affected by workload.

These results suggest that acoustical measures of the voice may reflect the increased effort mobilized by operators to perform a task when the demands of the task increase. However, the effect is obscured by variability which may be caused by many factors. For example, pilots may have learned to limit the inflections in their voices while flying an airplane. If so, they may have voluntarily, and unconsciously, limited any changes in the frequency of their voices. Another source of variability is the nature of the speech collected in these studies. The laboratory study required one short sentence, while the simulator task required much more lengthy spoken messages to air traffic controllers. The lengths of the utterances may have affected the breathing patterns of the pilots, which in turn may have affected the acoustical measures. There are many possible hypotheses as to why workload did not always affect acoustical measures in the present studies and those of others (e.g., Shipp, Brenner, & Doherty, 1986; Williams & Stevens, 1981).

The results for the drop-offs in the acoustical measures over time in the laboratory and simulator tasks suggest that the pilots' voices lost energy over the course of about two thirds of the runs. There were differences among the pilots as to whether the greater part of the drop-offs occurred early or late in the runs in the simulator. When the correlation between the acoustic measures and time were calculated, there were no statistically significant positive correlations, but several significant negative correlations, suggesting a reduction in the acoustical measures over time in both the simulator and laboratory tasks. Most of the negative correlations that were significant at the .02 level occurred in the high workload condition in both tasks. This result is in accord with the earlier study (Schneider, Alpert, & O'Donnell, 1989), which suggested that increased task demands were associated with more rapid loss of energy in the voice over the course of many utterances.

However, the results of the laboratory task did not predict the results of the simulator task. For example, most of the significant drop-offs in the acoustical measures in the simulator task were for subjects 1 and 2. However, in the laboratory task, all three subjects had significant drop-offs. Also, no pilot displayed a significant drop-off in frequency in the laboratory task, while frequency did fall in the simulator task, particularly in the high workload condition, for subjects 1 and 2. Consequently, the present study did not succeed in finding a way to profile the way an operator's voice responds to task demands, and then apply that profile in the operator's actual setting.

All the subjects subjectively rated the high workload condition as more demanding than the low workload condition. Subject 2 reported the greatest difference between the two conditions. That result is not surprising, since he was the only subject to use manual throttle in the high workload condition. Subject 1 reported the least difference between the two workload conditions. However, the acoustical measures do not suggest that subject 2 displayed the largest drop-offs and subject 1 the least. The drop-offs were largest for subjects 1 and 2, least for subject 3. There was thus no match between subjective and voice measures of workload. This result might reflect differences among the pilots in the way they subjectively rated task demands.

The results suggest that voice measures of workload could play a role in assessing the demands placed by new technology on operators. However, that role is limited by the variability among operators, and even within a single operator, of the effects of task demands on the voice. It appears that acoustical measures of the voice may reflect the effort that the operator is devoting to a task, and the fatigue resulting from sustained effort. In this way, acoustical measures of the voice can measure workload only indirectly, by revealing the strategy that the operator is using to apportion effort to the tasks.

Voice recognition and synthesis technology is increasingly being incorporated into the flight deck. The technology promises to free the overloaded channels of the eyes and hands, by allowing the operator to control more aircraft functions through voice commands and auditory responses. As this technology is developed, it will be important to design the advanced flight decks in a manner that minimizes the demands on the pilots. Ways must be found to accurately measure these demands. Subjective ratings, and psychophysiological measures are often used to measure task demands, but they suffer from the same problem encountered in the present studies: the measures are influenced by many extraneous factors, and therefore are susceptible to large variability. Future research might explore the usefulness of multivariate measures of workload, in which voice is combined with subjective and psychophysiological measures, with the intent of improving the reliability of measurement.

The present series of studies would suggest, though, that many subjects should be used in any study to assess workload using voice measures. The task demands posed by identical equipment are likely to vary from operator to operator. By using many subjects, it can be possible to determine which equipment configuration is least taxing to the greatest number of operators.

When using voice measures to assess workload, both the mean value of the acoustical measure, and its change over time, should be considered. The mean values may reflect the neuromuscular response to workload-induced stress. The drop-off over time may reflect the fatigue caused by sustained effort. The drop-off will in turn lower the mean values. The present results suggest that one method for assessing the drop-off is with linear regression, i.e., the Pearson product moment correlation.

For example, it might be necessary to compare two equipment configurations in the advanced flight deck simulator at NASA Langley Research Center. A group of subjects might be required to operate the simulator twice, once with each configuration. The order of the runs could be counterbalanced across subjects. The number of subjects should be large enough to observe differences among subjects: at least 15 to 20 is suggested, since the variability among subjects could be large. Factors unrelated to workload should be controlled to the extent possible. In particular, subjects should be sufficiently familiar with the technology so that the effects of novelty and learning are minimized.

The first analyses would determine whether either configuration is associated with larger amplitude, frequency, peak duration or stress than the other. T tests, such as those described in the earlier report, could be used for the comparisons as well; however, the power of t tests would be limited by the high inter-subject variance likely. Nevertheless, it could be seen whether either configuration brought about an increase in acoustical measures for a substantial majority of the subjects.

The most revealing analyses, however, might concern the drop-offs in the acoustical measures over time. About two-thirds of the acoustical measures obtained using both configurations may suggest drop-offs over time. These drop-offs can be observed as negative product moment correlations when the acoustical measure is correlated with time. A substantial majority of statistically significant correlations might occur for one of the configurations; such a result might suggest that the configuration is more tiring to use. The result could be confirmed with analyses of variance. Main effects for time for one configuration, but not the other, might point to a difference in task demands.

It is presently straightforward to perform acoustical analyses. While the present series of studies used proprietary software at New York University Hospital, there are several voice analytic packages which run on personal computers now available (e.g., Hypersignal Workstation from Hyperception, Dallas, Texas). It is now simple to digitize the voice, using hardware for the personal computer such as the Texas Instruments TMS-320 voice processor. The recent availability of these tools should make research into the voice available to a wide range of laboratories.

Conclusions

This series of studies concerning voice measures of workload has led to the following conclusions:

1. Higher levels of workload tend to increase the mean frequency, amplitude, and syllable duration in many person's voices.

2. There is a great deal of variation among individuals, and among voice samples from a single individual, in frequency, amplitude, and syllable duration. This variance may explain why in the present work, and in previous work from other laboratories, the effect of workload upon the mean values for the acoustical characteristics did not reach statistical significance.

3. It was not possible to predict how a single operator's voice would respond to increased workload in a flight simulator by assessing how his voice responds to increased workload under controlled laboratory conditions.

4. The effects of workload upon the acoustical properties of the voice is best demonstrated by measuring the change in the voice over time. Higher workload conditions accelerate the rate at which frequency and amplitude diminish over time.

5. Drop-offs in frequency and amplitude can be statistically demonstrated by comparing voice samples late in a trial with samples from earlier in the trial. This method can also demonstrate the failure of the voice to regain old levels of amplitude and frequency after rest periods, another feature of high workload conditions.

6. Drop-offs in frequency and amplitude can be demonstrated also through regression analyses. A negative slope over time suggests a drop-off.

7. Increased mean frequency and amplitude may reflect heightened effort devoted to a task. Faster drop-offs in frequency and amplitude may reflect the fatigue resulting from sustained effort. In this way, the acoustical parameters of the voice may reveal the strategy that an operator uses for allocating effort during demanding situations.

References

Cannings, R., Borland, R.G., Hill, L.E., & Nicholson, A.N. (1979) Voice analysis and workload during the letdown, approach and landing. Paper presented to the Aerospace Medical Association, Washington, DC.

Rosvold, H.E., Mirsky, A.F., Sarason, I., et al. (1956) A continuous performance test of brain damage. Journal of Consulting Psychology, 20, 343-350.

Schneider, S.J., Alpert, M., O'Donnell, R. (1989) Voice measures of workload in the advanced flight deck. NASA Langley Research Center, Hampton, VA: Contractors Report 4249.

Shipp, T., Brenner, M., & Doherty, E.T. (1986) Vocal indicators or psychological stress induced by task loading. U.S. Air Force Report USAFSAM-TSQ-86-3.

Williams, C.E. & Stevens, K.N. (1981) Vocal correlates of emotional states. In J.K. Darby (Ed.) Speech evaluation in psychiatry. New York: Grune & Stratton.

Table 1

Mean Values for Four Acoustical Parameters
Subject in Study 1

	Workload Level			
	Baseline	Low	Moderate	High
Peak Duration (csec)	23.95	26.27	24.86	28.21
Amplitude (cbel)	15.56	16.70	16.38	18.53
Frequency (Hz)	87.83	95.43	99.35	113.88
Stress	15.08	15.57	15.50	16.99

Table 2

Mean Values for Four Acoustical Parameters
Test Pilot 1

	Workload level		
	Low	Moderate	High
Peak Duration (csec)	21.05	21.16	20.71
Amplitude (cbel)	18.91	18.88	18.94
Frequency (Hz)	72.06	71.67	73.04
Stress	16.46	16.45	16.46

Table 3

Mean Values for Four Acoustical Parameters
Test Pilot 2

	Workload level		
	Low	Moderate	High
Peak Duration (csec)	27.93	27.27	26.75
Amplitude (cbel)	16.12	17.85	17.74
Frequency (Hz)	97.51	99.05	98.12
Stress	17.95	18.00	17.94

Table 4

Mean Values for Four Acoustical Parameters
Test Pilot 3

	Workload level		
	Low	Moderate	High
Peak Duration (csec)	24.19	24.77	24.11
Amplitude (cbel)	15.45	15.51	14.92
Frequency (Hz)	107.72	108.78	107.87
Stress	17.74	17.81	17.63

Table 5
Pearson Product Moment Correlations
Between Time and
Amplitude

Subject 1

run	baseline	low	medium	high
1	-.43	-.81***	.01	-.77***
2		.38	-.60+	-.36
3		-.61+	-.58+	-.26

Subject 2

run	baseline	low	medium	high
1	.06	-.13	-.67*	.32
2		-.46	-.75**	-.76***
3		-.33	-.39	-.73**

Subject 3

run	baseline	low	medium	high
1	-.32	-.61+	.09	-.60+
2		-.12	.13	-.72**
3		-.16	-.25	

+ $p < .10$
* $p < .05$
** $p < .02$
*** $p < .01$

Table 6
Pearson Product Moment Correlations
Between Time and
Frequency

Subject 1

run	baseline	low	medium	high
1	-.40	-.02	.02	.20
2		-.66*	-.29	-.28
3		-.41	-.42	.38

Subject 2

run	baseline	low	medium	high
1	.11	-.38	-.01	-.55+
2		.39	-.34	-.53
3		.30	-.30	-.30

Subject 3

run	baseline	low	medium	high
1	-.10	.10	-.34	-.43
2		.15	.10	.14
3		-.13	.06	

+ $p < .10$
* $p < .05$
** $p < .02$
*** $p < .01$

Table 7
Pearson Product Moment Correlations
Between Time and
Peak Duration

Subject 1

run	baseline	low	medium	high
1	-.41	-.61+	.38	-.29
2		-.17	-.33	-.20
3		-.17	.12	.36

Subject 2

run	baseline	low	medium	high
1	.11	-.50	-.58+	.10
2		.29	.17	.02
3		-.04	-.18	-.24

Subject 3

run	baseline	low	medium	high
1	.59+	-.58+	.41	-.51
2		.00	.03	-.31
3		.24	-.18	

+ $p < .10$
* $p < .05$
** $p < .02$
*** $p < .01$

Table 8
Pearson Product Moment Correlations
Between Time and
Stress (emphasis)

Subject 1

run	baseline	low	medium	high
1	-.57+	-.67*	.18	-.30
2		-.52	-.56+	-.39
3		-.61+	-.58+	.32

Subject 2

run	baseline	low	medium	high
1	.15	-.54+	-.51	-.28
2		.15	-.45	-.82***
3		-.24	-.62+	-.42

Subject 3

run	baseline	low	medium	high
1	-.05	-.36	-.18	-.72**
2		.12	.08	-.33
3		-.05	-.13	

+ $p < .10$
* $p < .05$
** $p < .02$
*** $p < .01$

Table 9
Test Pilots' Subjective Ratings of Difficulty in the Simulator Task

	First Half	Second Half	Overall
Subject 1			
low workload	1.23	1.58	1.32
high workload	1.76	1.84	1.80
Subject 2			
low workload	1.10	1.70	1.40
high workload	2.39	4.20	3.39
Subject 3			
low workload	1.60	2.55	2.08
high workload	2.65	3.65	3.15

Table 10

Mean Values for Four Acoustical Parameters, Simulator Task

	Subject 1 Workload			Subject 2 Workload			Subject 3 Workload		
	Low		High	Low		High	Low		High
Peak Duration (csec)	24.01	*	24.75	25.17		24.43	24.64	*	24.74
Amplitude (cbel)	20.66	*	20.94	12.17	*	12.76	18.02	*	18.61
Frequency (Hz)	91.47		88.61	117.96		116.18	118.29		107.74
Stress	16.90	*	16.97	17.54		17.35	17.39	*	17.47

Note--entries are the mean values for the combined first six and last six voice samples collected in the simulator, averaged across the two runs in each workload condition. Asterisks denote where the values for the high workload condition were greater than the values for the low workload condition.

Table 11
Changes in the Acoustical Measures
Simulator Task, Subject 1

utterance	peak duration (csec) workload low		amplitude (cbel) workload low		frequency (Hz) workload low		stress workload low	
	low	high	low	high	low	high	low	high
A) first 3	24.33	25.73	21.53	23.09	96.32	93.01	17.17	17.49
B) 10,11,12	21.49	24.00	18.95	20.36	88.82	89.78	16.42	16.80
C) last 3	25.98	23.73	21.35	20.16	89.73	86.54	17.13	16.75
A minus B	2.84	1.73	2.58	*2.73	7.50	3.23	.75	.69
B minus C	-4.49	* .27	-2.40	* .20	-.91	* 3.24	-.71	* .05
A minus C	-1.65	*2.00	.18	*2.93	6.59	6.47	.04	* .74

Note--the asterisks in Tables 11, 12, and 13 show where a drop-off in an acoustical measure was greater in the high workload condition than it was in the low workload condition.

Table 12
Changes in the Acoustical Measures
Simulator Task, Subject 2

utterance	peak duration (csec) workload		amplitude (cbel) workload		frequency (Hz) workload		stress workload	
	low	high	low	high	low	high	low	high
A) first 3	24.69	23.44	13.55	12.87	121.65	117.79	17.89	17.59
B) 10,11,12	26.16	23.81	12.89	10.06	120.64	112.55	17.80	16.98
C) last 3	22.45	22.03	9.27	12.71	105.98	114.93	16.60	17.55
A minus B	-1.47	-.37	.66	*2.81	1.01	* 5.24	.09	* .31
B minus C	3.71	1.78	3.62	-2.65	14.66	-2.38	1.20	-.57
A minus C	2.24	1.41	4.28	.16	15.67	2.86	1.29	-.26

Note--the asterisks in Tables 11, 12, and 13 show where a drop-off in an acoustical measure was greater in the high workload condition than it was in the low workload condition.

Table 13
Changes in the Acoustical Measures
Simulator Task, Subject 3

utterance	peak duration (csec) workload low	peak duration (csec) workload high	amplitude (cbel) workload low	amplitude (cbel) workload high	frequency (Hz) workload low	frequency (Hz) workload high	stress workload low	stress workload high
A) first 3	23.86	25.86	17.99	18.93	114.18	107.98	17.24	17.61
B) 10,11,12	23.99	22.54	17.65	17.08	117.30	101.43	17.25	16.95
C) last 3	25.14	26.72	18.67	19.15	121.14	113.43	17.62	17.83
A minus B	-.13	*3.32	.34	*1.85	-3.12	* 6.55	-.01	* .66
B minus C	-1.15	-4.18	-1.02	-2.07	-3.84	-12.00	-.37	-.88
A minus C	-1.28	-.86	-.68	-.22	-6.96	-5.46	-.38	-.22

Note--the asterisks in Tables 11, 12, and 13 show where a drop-off in an acoustical measure was greater in the high workload condition than it was in the low workload condition.

Table 14

Correlations between time and acoustical measures in the simulator task: Amplitude

subject	run	workload low	workload high
1	1	-.24	-.61*
1	2	-.02	-.38
2	1	-.59*	-.50+
2	2	-.38	.48
3	1	.38	.33
3	2	.23	-.10

+ $p < .10$
* $p < .05$
** $p < .02$
*** $p < .01$

Table 15

Correlations between time and acoustical measures in the simulator task: Frequency

subject	run	workload low	workload high
1	1	-.60*	.05
1	2	-.18	-.77***
2	1	-.59*	-.78***
2	2	-.04	.41
3	1	.38	.47
3	2	.33	-.09

+ p < .10
* p < .05
** p < .02
*** p < .01

Table 16

Correlations between time and acoustical measures in the simulator task: Peak duration

subject	run	workload	
		low	high
1	1	-.01	-.38
1	2	.17	-.06
2	1	-.10	-.09
2	2	-.13	-.24
3	1	.25	.26
3	2	-.15	-.14

+ $p < .10$
* $p < .05$
** $p < .02$
*** $p < .01$

Table 17

Correlations between time and acoustical measures in the simulator task: Stress

subject	run	workload low	workload high
1	1	-.32	-.53+
1	2	-.01	-.49+
2	1	-.57+	-.70**
2	2	-.19	-.31
3	1	.51	.40
3	2	.24	-.08

+ $p < .10$
* $p < .05$
** $p < .02$
*** $p < .01$

MEAN AMPLITUDE

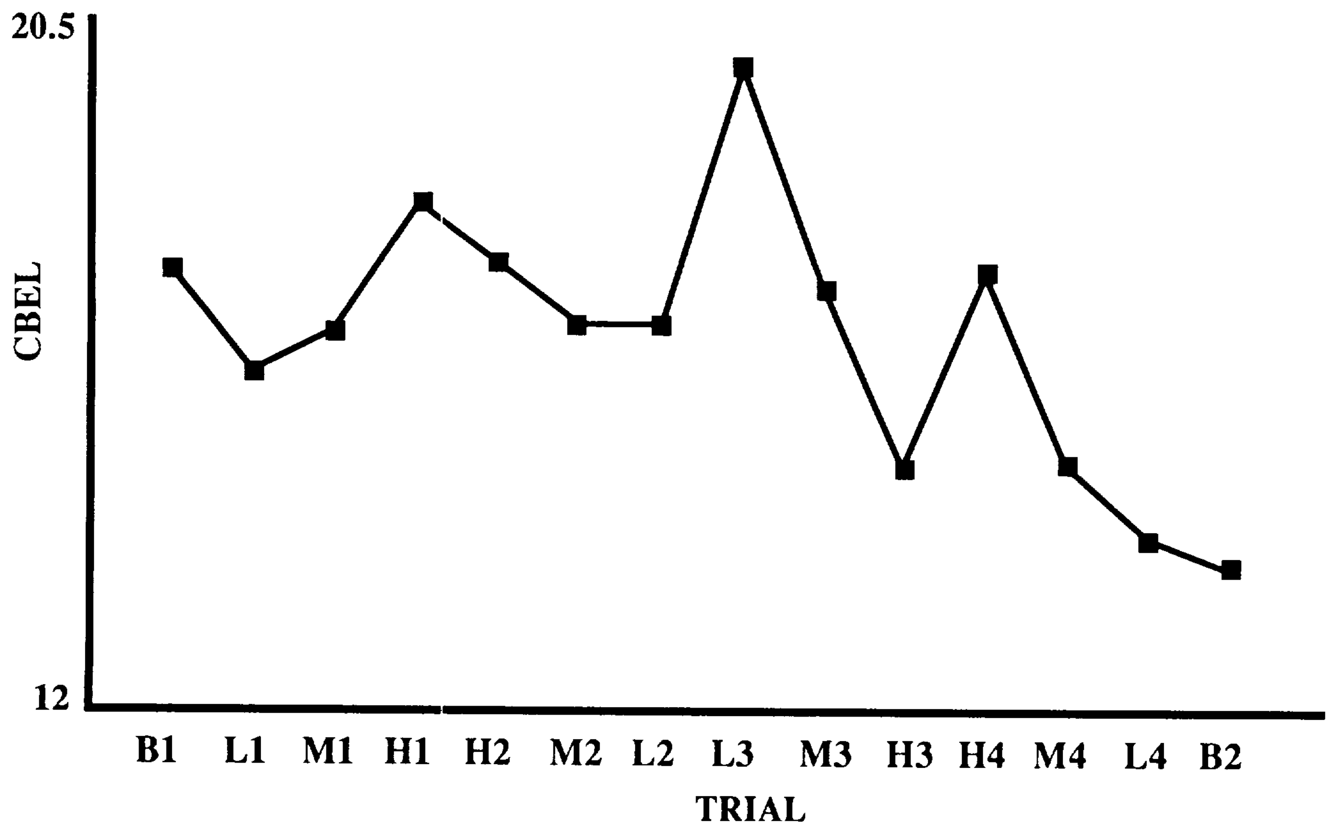

FIGURE 1

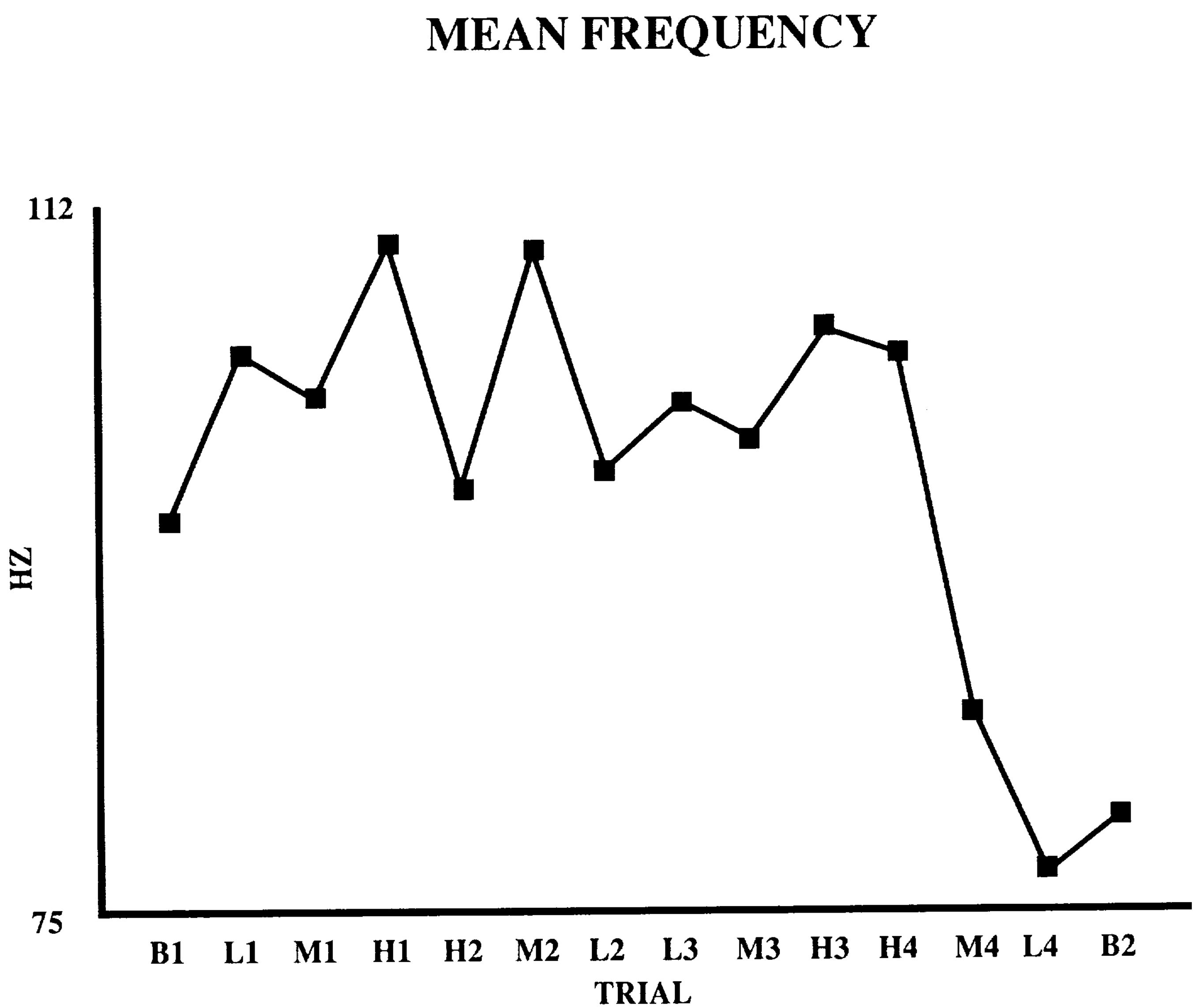
MEAN FREQUENCY
112
75
HZ
B1 L1 M1 H1 H2 M2 L2 L3 M3 H3 H4 M4 L4 B2
TRIAL

FIGURE 2

NASA National Aeronautics and Space Administration

Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NASA CR-4258		

4. Title and Subtitle	5. Report Date
Voice Measures of Workload in the Advanced Flight Deck: Additional Studies	November 1989
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
Sid J. Schneider and Murray Alpert	
	10. Work Unit No. 505-67-11-01
9. Performing Organization Name and Address Behavioral Health Systems, Inc. P. O. Box 547 Ossining, NY 10562	11. Contract or Grant No. NAS1-18278
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Langley Research Center Hampton, VA 23665-5225	13. Type of Report and Period Covered Contractor Report
	14. Sponsoring Agency Code

15. Supplementary Notes

Langley Technical Monitor: Alan T. Pope
Final Report

16. Abstract

These studies investigated acoustical analysis of the voice as a measure of workload in individual operators. In the first study, voice samples were recorded from a single operator during high, medium, and low workload conditions. Mean amplitude, frequency, syllable duration, and emphasis all tended to increase as workload increased. In the second study, NASA test pilots performed a laboratory task, and used a flight simulator, under differing workload conditions. For two of the pilots, high workload in the simulator brought about greater amplitude, syllable duration, and emphasis. In both the laboratory and simulator tasks, high workload tended to be associated with more statistically significant drop-offs in the acoustical measures than were lower workload levels. There was a great deal of intra-subject variability in the acoustical measures. The results suggest that in individual operators, increased workload might be revealed by high initial amplitude and frequency, followed by rapid drop-offs over time.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Voice Workload Acoustical measurement	Unclassified - Unlimited Subject Category 54

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of pages	22. Price
Unclassified	Unclassified	40	A03

NASA FORM 1626 OCT 86 NASA-Langley, 1989

For sale by the National Technical Information Service, Springfield, Virginia 22161-2171