Acoustic and perceptual indicators of emotional stress[a)]

L. A. Streeter
Bell Laboratories, Crawford Hill Laboratory, Holmdel, New Jersey 07733

N. H. Macdonald
Bell Laboratories, 6 Corporate Place, Piscataway, New Jersey 08854

W. Apple and R. M. Krauss
Columbia University, New York, New York 10027

K. M. Galotti
University of Pennsylvania, Philadelphia, Pennsylvania 19104

(Received 12 March 1982; accepted for publication 14 December 1982)

Tape recordings of telephone conversations of Consolidated Edison's system operator (SO) and his immediate superior (CSO), beginning an hour before the 1977 New York blackout, were analyzed for indications of psychological stress. (SO was responsible for monitoring and switching power loads within the Con Ed network.) Utterances from the two individuals were analyzed to yield several pitch and amplitude statistics. To assess the perceptual correlates of stress, four groups of listeners used a seven-point scale to rate the stress of SO and CSO from either randomized vocal utterances or transcripts of the randomized utterances. Results indicated that whereas CSO's vocal pitch increased significantly with increased situational stress, SO's pitch decreased. Listener ratings of stress from the voice were positively related to average pitch. It appears that listeners' stereotype of psychological stress includes elevated pitch and amplitude levels, as well as their increased variability.

PACS numbers: 43.70.Gr, 43.70.Dn, 43.66.Hg

INTRODUCTION

Voice indications of psychological stress are perhaps the most commonly studied emotional phenomena in speech production. The probable cause of this popularity, apart from certain obvious practical applications such as automatic lie detection, is the belief that stress sets into motion certain involuntary physiological responses that result in predictable changes in vocal parameters, such as pitch or amplitude level. Thus the assumed involuntary nature of the stress reaction leads to an expectation of stable acoustic consequences. Previous data, however, on this problem have been limited in quantity and scope. These data are from either: (1) catastrophic occurrences, such as recordings of pilots speaking directly prior to a fatal crash or the classic Hindenburg radio announcement (Williams and Stevens, 1969, 1972) or (2) laboratory studies in which stress is artificially induced. Examples of the latter include subjects attempting to deceive interviewers (Friedhoff *et al.*, 1964; Ekman *et al.*, 1976; Streeter *et al.*, 1977) or subjects performing a complex task under increasing time pressure (Hecker *et al.*, 1968).

Collectively, the results from the above studies are inconclusive. Naturalistic data indicate that vocal pitch (more precisely, fundamental frequency) rises in extremely stressful situations, but unfortunately the total amount of speech data has been limited as has been the number of voices sampled (five in Williams and Stevens, 1969 study). On the other hand, the consequences of experimentally induced arousal on fundamental frequency (F_0) are varied. In an excellent review of vocal indicators of emotional state, Scherer (1979) tabulated the average rise in F_0 observed in five laboratory studies. (For additional reviews of the vocal stress and emotional state literature, see Scherer, 1981 and Williams and Stevens, 1981.) In three studies there were negligible, nonsignificant increases in F_0 under conditions of arousal (Bonner, 1943; Rubenstein, 1966; Hecker *et al.*, 1968). In the remaining two studies the F_0 rise was statistically significant (Ekman *et al.*, 1976; Streeter *et al.*, 1977), but much smaller than observed in the natural contexts. A possible explanation for the apparent lack of consistency between results from naturalistic and laboratory studies is that the degree of arousal subjects experience in the contrived laboratory situations is minimal, thereby attenuating any vocal stress reaction.

In one laboratory study (Streeter *et al.*, 1977) the arousal variable was manipulated directly. In that study subjects attempted to deceive an interviewer. In one condition subjects were told that the ability to deceive successfully was correlated with general intelligence, and that their interview was to be videotaped and evaluated by a team of psychiatrists. This instruction presumably increased their motivation to lie and hence, indirectly increased arousal. The other group of subjects did not receive this instruction. The effect of these arousal instructions on average pitch level was marginally significant; there was a greater pitch difference between the truth-telling and lying conditions for aroused subjects compared to nonaroused subjects. (The average increase in F_0 was 6 Hz for aroused and 0.6 Hz for unaroused subjects.) Thus there is some indication that arousal increases F_0.

[a)] Based on a paper presented at the 96th Meeting of the Acoustical Society of America, Honolulu, Hawaii, 1978.

While it is possible that the null results obtained in some of the previous studies were due to insufficient experimental manipulations of arousal, it is by no means clear that stress necessarily affects vocal parameters such as pitch level in a unidirectional fashion. Even considering the two experimental studies (Ekman *et al.*, 1976; Streeter *et al.*, 1977) in which positive results were obtained, a sizable proportion (about 25%) of subjects in each study failed to show F_0 increases in the stressed condition. Similarly, Hecker *et al.* observed that while F_0 appeared to *change* under conditions of stress, individuals differed in the direction of the change. Of course, it may be that these reversals are simply due to normal statistical fluctuation.

The above discussion has focused exclusively on average pitch changes under stress, primarily because pitch has been the most frequently measured variable in previous research. While measures such as pitch variability, pitch range, total intensity, and energy distribution have been examined, no single study has examined all these measures. Again, the findings with regard to these variables have been ambiguous.

Thus the data collected to date appear to anchor the stressfulness continuum. The data of this paper come from a situation somewhere between the two anchor points—catastrophic occurrences and typical laboratory situations. In our study, neither of the two people whose voices we analyzed was in physical danger; nonetheless, for both, the situation as it developed became an acutely stressful one.

On 13 July 1977 failure of Consolidated Edison's electric power distribution system resulted in the complete blackout of Manhattan and Westchester County. Consolidated Edison (Con Ed) is the sole electric power company serving New York City. The 1977 blackout is considered to be the worst power failure in the Northeast since 1965. Causes of the blackout reportedly included: severe weather conditions, malfunctioning equipment, and human error (Federal Energy Regulatory Commission Report, 1978).

As prescribed by law, Con Ed tape-recorded telephone conversations of certain personnel in the main control center. In subsequent investigations by state, local, and federal commissions into causes of the blackout, recordings of relevant telephone conversations were supoenaed, and consequently brought into the public domain. It is the conversations of the Con Ed system operator beginning approximately 1h before the blackout and ending with the total blackout of New York that comprise our data base. The system operator had responsibility for monitoring and switching power loads within the Con Ed network. His duties included maintaining a balance between power generation and load, and during emergency situations ordering voltage reductions and initiating load-reduction procedures (Federal Energy Regulatory Commission Report, 1978).

Thus, one voice analyzed was the Con Ed system operator (SO) and the other was the chief system operator (CSO), the system operator's immediate superior who communicated by telephone from his home. Their conversation began about 20 min before the actual blackout and ended with the blackout. Both voices were analyzed for acoustic indications of stress.

In addition to acoustic analyses, we examined perceptual aspects of psychological stress (i.e., acoustic cues naive listeners use to infer that a speaker is aroused or relaxed). To this end, we presented randomized sequences of utterances from these conversations to listeners and asked them to judge the psychological stress of the speaker. A different group of subjects received the same randomized sequence of utterances, presented in a written transcript; these subjects also judged psychological stress. We were interested in determining: (1) whether there are reliable acoustic indicators of psychological stress, (2) what, if any, the perceptual indicators of psychological stress are, and (3) to what extent the listeners' stress ratings correctly mirror the actual degree of situational stress.

One of the most attractive features of our data was that situational stress presumably increased throughout this hour, which allowed for more than a binary classification of stress. This aspect differs from the other naturalistic data cited above and also from the previous laboratory studies. In addition, the amount of spoken material analyzed for each individual was large in comparison to previous studies.

I. METHOD

An audio tape of the system operator's conversation beginning at 8:36 p.m., and ending at 9:31 p.m. was digitized in a single session at a sampling rate of 10 000 samples/s. The utterances of the system operator (SO) and his superior (CSO) were excised using a waveform editing program implemented on a DDP-224 computer (Nakatani, 1977). Usually an "utterance" corresponded to a single turn-take in the conversation. However, there were cases in which a single turn-take consisted of two "sentences" separated by a lengthy pause. In these cases the single turn-take was divided into two utterances. There were 303 utterances for SO and 138 for CSO.

The transcript contained in the Federal Energy Regulatory Commission report (1978) served as the initial transcript. Minor corrections to this basic transcription were made by the first two authors by listening repeatedly to the audio tapes.

A. Acoustic, linguistic, and perceptual measurements

Each utterance as described above was measured in terms of the pitch, amplitude, and duration statistics. All utterances were submitted to a linear predictive coding analysis (Atal and Hanauer, 1971), which analyzes the speech waveform and produces 14 parameter values at 10 ms intervals (henceforth referred to as a sample). These include: fundamental frequency, speech amplitude, and 12 area function parameters, representing spectral characteristics of the speech.

Since the original tape recordings of the conversations contained noise, it was sometimes difficult to track F_0. Thus it was necessary to eliminate potentially spurious pitch values.[1] The resulting trimmed data for each *utterance* were measured in terms of: average F_0, maximum F_0, minimum F_0, and pitch variability (SD of F_0). For the amplitude measurements we considered amplitude values for only the voiced samples that met the above inclusion criteria. The amplitude statistics were the same as the pitch statistics: mean, maximum, minimum, and standard deviation of rms amplitude for each utterance. The timing measurements included three duration measurements: (a) total duration of each utterance, measured from the waveform display, (b) the number of syllables in each utterance, obtained from the written transcript, (c) the number of words in each utterance, also obtained from the transcript, and two speech rate statistics: (d) syllables/s and (e) words/s. (Since syllables/s and words/s were so highly correlated, subsequent tables show only one measure—words/s.)

To assess the perceptual correlates of psychological stress, groups of listeners rated the stress of either SO or CSO from either an audio tape of randomized utterances or a transcript of randomized utterances. All subjects judged speakers' stress using a seven-point rating scale with "one" indicating "not at all stressed" and "seven" indicating "extremely stressed." Subjects were told that the utterances had been collected beginning about an hour before the New York blackout, but that the utterances had been placed in random order, so they did not necessarily reflect the sequence of events that evening. The subjects were asked to rate the amount of stress the speaker experienced. There were four different rating groups: (1) SO audio (13 paid students from Columbia University), (2) SO transcript (13 paid college students from the Murray Hill, NJ area), (3) CSO audio (13 paid students from Columbia), and (4) CSO transcript (15 paid college students from the Murray Hill area). The random order of utterances for groups (1) and (2) and groups (3) and (4) was identical. In the audio condition each utterance was preceded by a tone with individual utterances separated by 5 s of silence. The transcript ratings were self-paced.

TABLE I. Summary statistics across all utterances for the System Operator (SO) and the Chief System Operator (CSO).

	SO $N = 303$		CSO $N = 138$	
Pitch variables (Hz)	Mean	SD (between utterances)	Mean	SD (between utterances)
Average F_0	139.8	20.8	150.1	14.5
SD F_0 (within utterance)	22.4	10.9	22.1	9.8
Maximum F_0	203.2	58.6	202.5	42.4
Amplitude variables (rms)[a]				
Average amplitude	566.6	85.0	277.0	83.6
SD amplitude (within utterance)	203.6	45.1	137.9	48.1
Maximum amplitude	872.2	80.8	558.3	165.4
Duration variables (s)				
Length	1.6	1.8	1.6	1.4
No. of words	6.5	6.5	7.7	6.2
Words/s	4.4	1.6	5.1	1.7
Ratings of stress (Seven-point scale)				
From audio	4.2	0.8	4.2	1.0
From transcript	3.9	1.1	4.4	1.0

[a] rms, or root mean square, is related to the more common measure, decibels, by the following formula: $dB = 20 \log_{10} rms$.

II. RESULTS

Table I shows the overall averages and standard deviations (SD) for each variable separately for SO and CSO.

As a first analysis, we assumed that situational stress increased monotonically throughout the sample. The utterances were ordered from first to last, and this ordinal measure of "time under stress" (utterance's serial position) was correlated with each of the dependent measures. Table II shows the correlations between serial position and the dependent measures, separately for SO and CSO. Notice that the pattern of correlations is different for the two individuals; whereas SO's pitch level and maximum amplitude decreased with time, the reverse was true for CSO. Also, over time the pitch contours for CSO's utterances became more variable with higher and higher pitch excursions, however, there were no comparable results for SO's voice. CSO also spoke more slowly with increasing time, while SO's rate of speech did not change measurably with time. In short, these two individuals show different acoustical manifestations of increasing situational stress.

To justify the use of an utterance's serial position as a measure, we wanted some external validation of the assumption that situational stress, in fact, increased throughout the hour. We referred to two external sources to obtain corroboration: the Federal Energy Commission's report (1978) and two power controllers from a suburban electric company. The power controllers were given copies of the transcript to read a few days prior to a telephone interview. These controllers were familiar with the technical content of the transcript as their jobs were similar to SO's and CSO's

TABLE II. Correlation between utterance serial position and dependent measures for SO and CSO.

	SO	CSO	p of difference
Pitch variables			
Average F_0	-0.16^b	0.25^c	c
SD F_0 (within utterance)	-0.05	0.22^b	a
Maximum F_0	-0.04	0.24^b	a
Amplitude variables			
Average amplitude	-0.01	0.24^c	a
SD amplitude (within utterance)	-0.11^a	0.26^c	c
Maximum amplitude	-0.19^c	0.24^b	c
Duration variables			
Length (s)	-0.10^a	0.22^b	b
No. of words	-0.10^a	0.13	a
Rate (words/s)	0.01	-0.14^a	
Ratings of stress (seven-point scale)			
From audio	-0.14^b	0.23^b	c
From transcript	0.00	0.08	

[a] $p < 0.05$.
[b] $p < 0.01$.
[c] $p < 0.001$.

albeit for a much smaller geographical area.

The controllers rated each of SO's conversations in chronological order using a seven-point scale for the amount of situational stress they believed SO to be under. The controllers were told to rate the severity of the situation with reference to the events that had preceded. The long conversation between SO and CSO was treated differently from the conversations that SO had with other power personnel. For the long conversation, the controllers indicated places in the transcript where they would either increase or decrease their stress ratings. The controllers' ratings corroborated our assumption of monotonically increasing situational stress throughout this hour.

The second validation of increasing situational stress comes from the Federal Energy Regulatory Commission's (FEC) final report. In that report the sequence of events that led to the total collapse of the system are outlined. The report states that "From 8:37 p.m. to 9:30 p.m. on July 13th, a series of seven incidents occurrred, each one in turn either decreasing Con Edison's capability to import power through interconnections, or decreasing its own system generation, or both (p. 19)." According to the report, at the beginning of the hour under investigation, the Con Ed network was "in normal stable operation." By the end of the hour the entire city was without power. The series of incidents enumerated in the report served as the basis for dividing the data into sequential stress periods. We divided the hour into five categories, described below. However, subsequent analyses deal with only three of the periods because the first (before any of the seven incidents occurred) and last period contained too few utterances (four and five, respectively). The three categories analyzed below are referred to as the first, second, and third stress periods.

The first stress period began when lightning struck a major transmission line. However, after this initial mishap conditions were considered to be stable. The first stress period lasted 19 min and included 59 utterances by SO and none by CSO. The second stress period began with a second lightning bolt, which resulted in isolating yet another large transmission line from the rest of the network. According to the FEC report, the system was in a serious emergency state after this second incident, since there was severe overloading of the remaining "vital" interconnections. The second stress period lasted 23 min and included 68 of SO's utterances and 85 of CSO's. In the third period, four interconnection lines were lost in rapid succession, resulting in a 30% generation deficiency. This large deficiency then caused the collapse of the system. The duration of the final stress period was 11 min and included 67 of SO's utterances and 52 of CSO's utterances.

Table III presents averages of each dependent variable for the three stress periods for SO and for the final two stress periods for CSO. The rightmost column indicates the significance level of mean differences between the first and third periods for each variable for SO and the second and third periods for CSO. Once again, the pattern of means shows that the acoustic variables changed differently for the two men; whereas CSO's pitch and amplitude increased over time, SO's pitch decreased and his amplitude changed little.

TABLE III. Statistics for each stress period.

	SO Stress period 1st $N=59$	2nd $N=168$	3rd $N=67$	1st–3rd $df=124$
Pitch variables (Hz)				
Average F_0	146	138	136	[b]
SD F_0 (within utterance)	24	22	22	
Maximum F_0	209	202	197	
Amplitude variables (rms)				
Average amplitude	574	560	578	
SD amplitude (within utterance)	216.5	201.1	197.7	
Maximum amplitude	909	864	862	
Duration				
Length (s)	1.6	1.8	1.1	[a]
No. of words	5.9	7.6	4.2	
Rate (words/s)	3.9	4.6	4.1	
Ratings of stress (seven-point scale)				
From audio	4.2	4.2	4.0	
From transcript	3.8	3.9	3.8	

	CSO Stress period 2nd $N=85$	3rd $N=52$	2nd–3rd $df=135$
Pitch variables (Hz)			
Average F_0	147	155	[c]
SD F_0 (within utterance)	20	26	[c]
Maximum F_0	193	218	[c]
Amplitude variables (rms)			
Average amplitude	255	313	[c]
SD amplitude (within utterance)	124.1	160.4	[c]
Maximum amplitude	515	629	[c]
Duration			
Length (s)	1.3	2.1	[b]
No. of words	6.9	9.0	
Rate (words/s)	5.3	4.8	
Ratings of stress (seven-point scale)			
From audio	4.0	4.5	[b]
From transcript	4.3	4.6	

[a] $p < 0.05$.
[b] $p < 0.01$.
[c] $p < 0.001$.

In addition, many of the duration variables differ in direction between SO and CSO. That these findings based on the "critical incidents" parallel those in Table II (based on correlations between acoustic variables and utterances' serial position), also increases our confidence that situational stress did, in fact, increase throughout the 1-h period. Thus we conclude that increasing situational stress *per se* does not result in a single pattern of acoustic consequences.

An additional feature of the data can be seen in Table IV, which shows SO's data from the second stress period

TABLE IV. SO while talking to CSO or talking to others in second stress period.

	CSO $N=107$ Mean	CSO SD	Others $N=61$ Mean	Others SD	p
Pitch variables (Hz)					
Average F_0	132.5	16.29	148.5	25.06	c
SD F_0 (within utterance)	21.0	9.93	23.2	11.09	
Max F_0	202.3	60.10	202.0	55.16	
Amplitude (rms)					
Average amplitude	553.4	73.76	570.8	101.09	
SD amp (within utterance)	195.3	37.94	211.2	54.25	a
Maximum amplitude	854.4	65.20	880.0	119.42	
Duration					
Length (s)	2.1	2.49	1.4	1.14	a
No. of words	7.9	8.25	7.1	6.09	
Rate (words/s)	4.2	1.67	5.3	1.73	c
Ratings of stress (seven-point scale)					
From audio	4.1	0.57	4.4	0.90	a
From transcript	4.0	1.01	3.9	1.05	

[a] $p<0.05$.
[b] $p<0.01$.
[c] $p<0.001$.

partitioned into those addressed to CSO and those addressed to others. (This was the only stress period in which SO conversed both with CSO and with others.) Again, the rightmost column indicates the significance level of the differences between the two speakers. There were significant differences between them for many of the acoustic variables. Unfortunately, interpreting these differences is problematic since time and addressee are confounded, i.e., conversations with CSO came towards the end of the third stress period. However, assuming roughly equivalent degrees of situational stress for the two during the third period, what explanations can be offered for the observed differences? First, SO may have been attempting to mask his degree of concern or emotional upset to CSO. Perhaps, SO tacitly knew that changes in certain vocal parameters can communicate an impression of a person "in control," and in an attempt to convey calmness, SO lowered his pitch, etc. While this explanation may be true, many of SO's statements indicate that he was attempting to communicate to CSO not only the gravity of the situation, but his own inability to cope with it adequately as well. In fact, it appears that SO was attempting to inform CSO of the situation quickly in order to transfer the decision-making burden to his superior. The government report also makes note of this transfer of decision-making:

> The Con Edison system operator managed to transfer decision making responsibility to his supervisor, who was in no position to respond adequately because he was at home and had limited knowledge of the conditions before or during the disturbances... It should be noted that the decision of the system operator to call his supervisor was purely discretionary. The operating procedures of Con Edision permit but do not require the system operator to contact his supervisor in times of system emergency (pp. 27–28).

If it were the case that the person making the decisions is most affected by situational stress, then we would expect to observe changes in acoustic variables for CSO's voice as a function of time. Indeed, this was the case as evidenced by the data presented above (see Tables II and III). Also, we would expect the *magnitude* of the shift in the acoustic variables to be greater for CSO than for SO. Again, the data appear to support this notion.

Table V shows the natural covariation among acoustic variables for SO and CSO. These data are provided primarily

TABLE V. Inter-correlations among the acoustic variables for SO and CSO. For SO: $r>0.094$, $p<0.05$. For CSO: $r>0.140$, $p<0.05$.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) Average F_0 (Hz)	...								
	...								
(2) SD F_0 (Hz)	0.53	...							
	0.50	...							
(3) Maximum F_0 (Hz)	0.50	0.75	...						
	0.61	0.77	...						
(4) Average amp (rms)	0.21	−0.07	−0.16	...					
	0.60	0.49	0.48	...					
(5) SD amplitude (rms)	0.19	0.07	−0.01	0.10	...				
	0.61	0.56	0.59	0.88	...				
(6) Maximum amp (rms)	0.31	0.17	0.17	0.42	0.42	...			
	0.55	0.53	0.58	0.85	0.94	...			
(7) Length (s)	−0.06	0.08	0.34	−0.30	−0.15	0.15	...		
	0.05	0.28	0.41	0.24	0.35	0.50	...		
(8) No. of words	0.00	0.12	0.39	−0.27	−0.21	0.19	0.88	...	
	0.01	0.19	0.34	0.20	0.29	0.46	0.89	...	
(9) Rate (words/s)	0.13	0.07	0.07	0.21	−0.19	0.05	−0.17	0.19	...
	0.14	−0.27	−0.22	−0.11	−0.16	−0.12	−0.20	0.13	...

for completeness, and can be compared to other measurements in the literature.

III. SUBJECTS' RATINGS OF STRESS

While the acoustic indicators of stress appear to be quite complex and varied, the perceptual correlates—the acoustic cues listeners use to infer that state in a speaker—are more regular.

Table VI shows for both SO and CSO the correlations between listeners' subjective ratings of stress level (on a seven-point scale) in the audio condition. Interrater reliabilities (calculated by the method of Winer, 1971, pp. 283–287) were 0.98 for subjects rating SO's utterances, and 0.82 for subjects rating CSO's utterances. The rightmost column indicates the significance level of those variables for which the correlations differed reliably between SO and CSO.

Listeners' stress ratings correlated significantly with all of the pitch variables for both individuals. However, for CSO listeners' judgments correlated with amplitude measures as well. Thus, the acoustic cues listeners used in inferring stress appear to be to some extent, speaker dependent. These differences in turn may be related to the productions of the two speakers. SO exhibited somewhat greater variability for the pitch variables than did CSO (see Table I), whereas CSO displayed greater amplitude variability. However, more explanation is needed to see that this is case. In the original recordings the microphone was located with SO's telephone; CSO's utterances were recorded from SO's telephone. Consequently, CSO's average amplitude was attenuated and the amplitude range was restricted in the recordings. Therefore, observing approximately the same amount of variability for average amplitude (SO = 85.0, CSO = 83.6 rms) in fact indicates that average amplitude was more variable for CSO. Thus it appears that listeners may take account of the variability of acoustic parameters such as pitch and amplitude in the speech of the person they are judging, and weigh more heavily those parameters with greater deviations from the average.

TABLE VI. Correlations between listeners' ratings of stress and acoustic variables for SO and CSO.

	SO	CSO	p of difference
Pitch variables			
Average F_0	0.48^c	0.22^b	b
SD F_0 (within utterance)	0.45^c	0.38^c	
Maximum F_0	0.43^c	0.47^c	
Amplitude variables			
Average amplitude	0.08	0.46^c	c
SD amplitude (within utterance)	− 0.02	0.48^c	c
Maximum amplitude	0.27^c	0.58^c	c
Duration variables			
Length	0.25^c	0.42^c	
No. of words	0.33^c	0.53^c	a
Words/s	0.14^b	0.15^a	
Perceptual			
Ratings from written transcript	0.48^c	0.06	c
Utterance serial position	− 0.48^c	0.23^b	

[a] $p < 0.05$.
[b] $p < 0.01$.
[c] $p < 0.001$.

Curiously, subjects who rated the degree of stress in the written transcript produced ratings which did not correlate significantly with the ratings of the listeners for CSO, but did correlate significantly for SO. Interrater reliabilities for subjects rating the written transcripts were 0.84: for SO, and for CSO's utterances 0.89.

While there appears to be a great deal of consistency in listeners' ratings of stress, are these ratings veridical, i.e., do they reproduce the ordering of events? In the case of SO, listeners' ratings of stress were significantly correlated with utterance serial position, but in the wrong direction. Of course, this has to be the case if listeners associate greater stress with higher pitch levels and greater pitch variability. This strategy produces the "correct" result for CSO, but not for SO, whose pitch level declined over the hour in question. Therefore, we conclude that listeners' stereotype of psychological stress includes elevated pitch and amplitude levels as well as increased variability of these two variables.

IV. DISCUSSION

Our answer to the initial question, namely, are there reliable and valid acoustic indicators of psychological stress, is a qualified no. With two individuals we observed two different and in many cases opposite patterns of acoustic variables as a result of what must have been increasing situational stress. Thus, to the extent that situational stress elevates psychological stress, our results argue against a single, stable stress reaction at least in terms of the acoustic variables measured. However, the confounded nature of the data base makes it difficult to support this conclusion with certainty. For instance, it is possible to argue that SO did not internalize the situational stress; that his transference of decision making exempted him from feeling and acting stressed. While this is a possibility, it does not seem probable. The nature of the data however, makes it impossible to eliminate this interpretation. It is also possible that SO could have adapted to the stress, so that the first measurement period encompassed the time SO experienced the greatest amount of stress. Alternatively, SO may have been in a stress overload situation which may differ from the normal stress reaction.

The *ad hoc* nature of the above explanations only serve to underscore the difficulties inherent in inferring a particular emotional state based soley on the external events. Most importantly however, it points to the dangers of relying on acoustic data to establish a person's internal state. Thus what these data do indicate is that a particular sequence of events which can be agreed is stressful does not necessarily result in predictable vocal behavior. On the other hand, what is predictable to a large degree is *listeners'* behavior. Listeners view certain vocal behaviors as indicative of particular emotional states. That is, listeners seem to refer to a vocal stereotype or a complex of cues that they normally associate with stress, which include elevated pitch and amplitude levels as well as greater variance of these cues. This vocal stereotype while consistent across listeners can be notably inaccurate.

ACKNOWLEDGMENT

We are grateful to the New York City Department of Investigations for furnishing us with a copy of the recording that we used for this study.

[1]By visual inspection of the pitch trackings for a number of utterances, it appeared that pitch errors were characterized by an unsystematic dispersion of pitch values (i.e., the absence of a smooth F_0 contour). There were additional errors attributable to pitch halving and doubling. No attempt was made to correct any of these errors; rather, based on the characteristics of the pitch trackings, these questionable data points were eliminated. All F_0 values lower than 80 Hz or greater than 350 were discarded. Also, we required adjacent pitch vlaues to be within 20 Hz of one another. (The removal of these suspicious pitch values did not appreciably affect average F_0 but did affect the pitch variance measures.)

Atal, B. S., and Hanauer, S. L. (**1971**). "Speech analysis and synthesis by linear prediction of the speech wave," J. Acoust. Soc. Am. **50**, 637–655.

Black, J. W. (**1961**). "Relationships among fundamental frequency, vocal sound pressure, and rate of speaking," Lang. Speech **4**, 196–199.

Bonner, R. (**1943**). "Changes in the speech pattern under emotional tension," Am. J. Psychol. **56**, 262–273.

Ekman, P., Friesen, W. V., and Scherer, K. R. (**1976**). "Body movement and voice pitch in deceptive interaction," Semiotica **16**, 23–27.

Federal Energy Regulatory Commission, U.S. Department of Energy, "The Con Edison power failure of July 13 and 14, 1977," Final staff report, June 1978.

Friedhoff, A. J., Alpert, M., and Kurtzberg, R. L. (**1964**). "Infracontent channels of vocal communication," Assoc. Res. Nerv. Ment. Dis. **42**, 414–423.

Hecker, M. H. L., Stevens, K. N., von Bismarck, G., and Williams, C. E. (**1968**). "Manifestations of task-induced stress in the acoustic speech signal," J. Acoust. Soc. Am. **44**, 993–1001.

Nakatani, L. H. (**1977**). "Computer-aided signal handling for speech research," J. Acoust. Soc. Am. **61**, 1057–1062.

Rubenstein, L. (**1966**). "Electro-acoustical measurement of vocal responses to limited stress," Behav. Res. Therapy **4**, 135–141.

Scherer, K. R. (**1979**). "Nonlinguistic vocal indicators of emotion and psychopathology," in *Emotions and Psychopathology*, edited by C. E. Izard (Plenum, New York), pp. 493–529.

Scherer, K. R. (**1981**). "Vocal indicators of stress," in *Speech Evaluation in Psychiatry*, edited by J. K. Darby (Grune–Stratton, New York).

Streeter, L. A., Krauss, R. M., Geller, V., Olson, C., and Apple, W. (**1977**). "Pitch changes during attempted deception," J. Personal. Soc. Psychol. **35**, 345–350.

Williams, C. E., and Stevens, K. N. (**1969**). "On determining the emotional state of pilots during flight: An exploratory study," Aerosp. Med. **40**, 1369–1372.

Williams, C. E., and Stevens, K. N. (**1972**). "Emotions and speech: some acoustical correlates," J. Acoust. Soc. Am. **52**, 1238–1249.

Williams, C. E., and Stevens, K. N. (**1981**). "Vocal correlates of emotional states," in *Speech Evaluation in Psychiatry*, edited by J. K. Darby (Grune–Stratton, New York).