ATTACHMENT 7

Reprints of two scientific articles addressing aspects of speech analysis in the Commonwealth of Independent States
(9 pages)

DOT/FAA/AM-96/10

Office of Aviation Medicine
Washington, D.C. 20591

Methods and Metrics of Voice Communications

Barbara G. Kanki, Editor
Ames Research Center
National Aeronautics and Space Administration
Moffett Field, CA 94035

O. Veronika Prinzo, Co-Editor
Civil Aeromedical Institute
Federal Aviation Administration
Oklahoma City, Oklahoma 73125

Proceedings of a workshop organized by the Federal Aviation Administration, National Aeronautics and Space Administration, and Department of Defense.
San Antonio, Texas, May 13-14, 1995.

March 1996

Final Report

This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

U.S. Department of Transportation

Federal Aviation Administration

SPEECH ANALYSIS IN RUSSIA

Malcolm Brenner, David Mayer, James Cash
National Transportation Safety Board

INTRODUCTION

This paper resulted from the May 1989 Agreement on Cooperation in Transportation Science and Technology between the United States and the former Soviet Union. As part of the original agreement, a subgroup for Aircraft Accident Investigation was formed. The National Transportation Safety Board (NTSB) and the GOSAVIANADZOR of the Soviet Union began cooperative technical exchanges of specialists and material related to accident investigation and prevention. Following the 1991 breakup of the former Soviet Union, the cooperative exchanges continued between the NTSB and the newly formed Interstate Aviation Committee (MAK) that represents the accident investigation authorities of the Commonwealth of Independent States (CIS). This paper resulted from a continuation of the cooperative work of the Accident Investigation Group.

There has been an exchange of papers and personal visits related to areas of scientific cooperation, exchanges of the sort that were not possible during the political climate that prevailed between our countries in most of the recent past.

In line with this effort, our agency provided information to our colleagues in the CIS concerning speech analysis work that was accomplished by our staff (Brenner, M., & Cash, J.R., 1991; Brenner, M., Doherty, E.T., & Shipp, T., 1994). In return, we received a remarkable letter from Alfred Belan, M.D., chief of the acoustics laboratory of the Interstate Aviation Committee in Moscow. The letter, written in broken English, claimed an ambitious program of speech analysis work of which we were completely unaware. The letter indicated that Dr. Belan was preparing a book in Russian describing observations made from the speech recordings of more than 300 airplane accidents. It should be noted that there are perhaps 30 airplane accident voice tapes discussed in English-language articles (Ruiz, R., Legros, C., & Guell, A., 1990). The letter, then, suggested a level of experience that was an order of magnitude greater than that of the entire scientific literature! Intrigued, we invited Belan to visit the United States for further discussions.

In February 1994, Dr. Belan spent a one-week visit at the NTSB headquarters in Washington, D.C. In addition to our staff, Barbara Kanki of NASA-Ames Research Center attended the meetings. The meetings consisted of both discussions and laboratory analysis of accident tapes.

Dr. Belan was a pleasant man in his late fifties, highly educated, who spoke little English but displayed a clever and charming sense of humor. Some of the credibility assigned to the Russian research came from the very favorable impression made by Dr. Belan himself, especially given the inherent language difficulties.

The information described in this paper is based on our meetings with Dr. Belan. This represents our best, albeit limited, understanding at the time of the Russian program.

Origin of the Russian Speech Analysis Program

The Russian speech effort began about 20 years ago and was centered in the Institute of Aviation Medicine. The work was inspired by the 1969 paper of American researchers Williams & Stevens (Williams, C.E., & Stevens, K.N., 1969). Early work from the Russian program was published in English (Simonov, P.V., & Frolov, M.V., 1973; Simonov, P.V., & Frolov, M.V., 1977). However, after the late 1970s, the work was no longer published outside Russia and it apparently took on something of a secret quality. Speech analysis was used to evaluate cosmonauts and pilots for fitness for duty in terms of both stress, fatigue, and other aeromedical qualities.

The program used simulator research, in some cases with test pilots as subjects, and also studied pilots and cosmonauts during real life aerospace situations. In the case of fatigue, for example, subjects performed in research projects for periods of 72 hours without sleep. Fatigue studies were made of cosmonauts in extended duty situations. In addition to research, systematic examination was made of aviation accident tapes from both military and civilian accidents.

Measures Used in the Russian Research

Dr. Belan referred to numerous speech measures used in Russian research. Although some were new to us, many were familiar from English language literature. What was striking about the Russian approach was its broadness, combining acoustic, phonetic, and communication information in a way that seemed original. What was also striking was the seriousness with which the measures were applied and the level of experience shown with the measures.

The Russian effort groups speech measures into 4 categories, which are evaluated for each speech sample. The categories are:

1) *acoustic measures*. These include fundamental frequency; fundamental frequency range; amplitude; and relative energy distributions among the formants. The last measure was of special interest, following from early work published by Russian researchers (Simonov & Frolov, 1973, 1977). At least some of these measures are extracted by automated techniques.
2) *timing measures*. These include speaking rate, and measures such as relative speaking/silence time and latency to respond.
3) *contour measures*. These relate to the relative shape of the speech energy waveform when plotted over time. An example would be whether the waveform is relatively flat or spiked.
4) *psycholinguistic measures*. These include phonetic measures such as changes in articulation of words. They also include measures of communication, such as whether communication is appropriate and effective given the ongoing conversation and the demands of the flight situation. One of the most interesting aspects of the Russian work is that it formally compares evidence based on the physical properties of speech with evidence based on the effectiveness of communication.

Proposed Standards

Based on his experience, Belan suggested general standards that apply to normal human speech. We have not seen such standards published and found them immediately practical in our work. We report them here for review by our colleagues.

For fundamental frequency, Belan suggested that a male speaker engaged in relaxed communication should display an average fundamental frequency between 80-130 Hz. The range should be higher, 95-145 Hz, in cockpit situations (perhaps because the speaker is compensating for background noise). Thus, if a pilot displays an average fundamental frequency that is higher than 145 Hz, regardless of the flight situation, it is abnormal and a sign that the pilot is very tense. (Belan noted, however, that intra-individual changes are more important than absolute changes on all speech measures).

For fundamental frequency range, Belan suggested that an average range of 45-75 Hz was normal in a relaxed situation. A range of 45-90 Hz was normal for a dynamic flight situation.

For speaking rate, Belan suggested an average rate of about 4.5 to 7.5 syllables per second as normal. A phrase might contain as few as 4-7 syllables, and in some cases as few as 2-3 syllables if the words were conversational, and still provide useful data for measuring speaking rate.

For segmenting statements, Belan suggested that a silent period of 300 msec be used to delineate the end of one statement and the beginning of another. This might represent an approximate minimum time necessary for a human speaker to shift thoughts.

An Example of Russian Work: Psychological Stress

As an example of Russian work, Belan described in detail some work on the speech effects of psychological stress. He provided a lecture on this topic, and demonstrated his thinking in a laboratory analysis of several accident tapes.

In general, the Russian work discusses 3 stages in the human response to psychological stress. These range on a continuum from a constructive response to absolute panic. The stages can be characterized as follows:

Stage 1. Belan described the first stage of stress as a working stress that improves performance, a constructive mobilization of attention and resources in reaction to an unusual event. The speaker is in control of speech, communications are accurate and there are no logical or semantic disturbances evident in speech. The pilot's performance in the cockpit shows no procedural errors. In acoustic and rate measures, this stage is characterized by an intra-individual increase of about 30% in fundamental frequency when compared to relaxed levels, an increase of about 10% in amplitude, and, perhaps, an increase of 5-10% in speaking rate.

Stage 2 . The second stage of stress was described as just strain. The pilot can still do the job and make decisions. Movements can become sharper but are still under control. The pilot does not make gross mistakes.

In the second stage of stress, speech is still adequate to the situation but emotional stress is clearly seen. Speech is fast, strained, brief, and accented. There may be a reduced latency to respond (such as the speaker's response beginning before the query is complete). Occasionally, phrases are not completed. Belan noted that there is a reduction of nonessential speech: the speaker "observes the purpose of communication." Speech may be repetitious as if to ensure that the recipient understands.

In Stage 2, the speaker's performance often displays hasty or premature actions. Intermediate procedural steps may be skipped, such as the omission of checklist items. The speaker appears to be trying to overtake the situation.

Stage 2 speech is characterized by an increase of 50-150% in fundamental frequency when compared to relaxed levels, an increase of 15-20% in amplitude, and, perhaps, an increase in speaking rate of more than 50%. Other signs of stress include an increase in fundamental frequency range and contour changes. Measures of pulse and respiration would show increases.

Stage 3. On top of all else, during Stage 3, the pilot cannot think straight. Sometimes he cannot speak clearly, leaves out letters, and repeats the same thing. Sometimes his answer is unrelated to the question. He is apparently thinking of something else. Belan says that speech is characterized by those things that dominate the speaker's thinking regardless of the situation. Standard operating procedures are not followed. There can be an occasional, stupor-like refusal to act (although this is rare).

In Stage 3, there is often incomplete articulation, with unvoiced syllables and words swallowed or not produced. There is poor word choice and improper grammar, and no attempt to correct speech errors. Fundamental frequency increases 100-200% over relaxed levels, amplitude increases 30-50%, and there can be large oscillations in rate including increases of 50-200%. Dr. Belan noted, however, that these changes may not apply to the highest levels of Stage 3. It is not unusual to see a sudden drop in fundamental frequency and hoarseness when the speaker faces imminent death.

Other Applications

Belan indicated that Russian work has examined fatigue and hypoxia effects on speech, areas in which there is no literature in English language journals. There is also work published in Russian on the physiology of physical effort and its effects on speech. These areas were discussed only briefly in our one-week meeting, however we received an impression that Russian work in these areas was as thoughtful as the work on psychological stress.

Future Directions

The Russian work appears to add significantly to previous work published in English language sources. It adds confidence that there may be characteristics of human speech that are cross-cultural and that will allow us to identify and quantify emotional responses. The leadership of the NTSB and MAK plan to continue the support of the cooperative exchanges of technical information and specialists in the field of accident investigation, and we anticipate further exchanges with the Russian program that can lead to a more involved cooperative work.

ACKNOWLEDGMENTS

We thank Eugenia Bernstein and the Russian Embassy for translation assistance during the visit of Dr. Belan. The opinions expressed in this paper are those of the authors and do not necessarily reflect the official position of the National Transportation Safety Board.

REFERENCES

Brenner, M., & Cash, J.R. (1991). Speech analysis as an index of alcohol intoxication — the Exxon Valdez accident. *Aviation, Space, and Environmental Medicine*, 62, 893-8.

Brenner, M., Doherty, E.T., & Shipp, T. (1994). Speech measures indicating workload demand. *Aviation, Space, and Environmental Medicine*, 65, 21-6.

Ruiz, R., Legros, C., & Guell, A. (1990). Voice analysis to predict the psychological or physical state of a speaker. *Aviation, Space, and Environmental Medicine*, 61, 266-71.

Simonov, P.V. & Frolov, M.V. (1973). Utilization of human voice for estimation of man's emotional stress and state of attention. *Aerospace Medicine*, 44, 256-8.

Simonov, P.V. & Frolov, M.V. (1977). Analysis of the human voice as a method of controlling emotional state: achievements and goals. *Aviation, Space, and Environmental Medicine*, 48, 23-5.

Williams, C.E. & Stevens, K.N. (1969). On determining the emotional state of pilots during flight: an exploratory study. *Aerospace Medicine*, 40, 1369-1372.

DEVELOPMENT OF A SPEECH ANALYSIS PROTOCOL FOR ACCIDENT INVESTIGATION

David L. Mayer, Malcolm Brenner, and James R. Cash
National Transportation Safety Board

INTRODUCTION

Accident Investigation

Evidence provided by voice recordings is often integral to the investigation of aviation accidents. These voice tapes may be recordings of radio traffic, or they may come from cockpit voice recorders (CVRs), which store the final 30 minutes of flight-deck sounds. These recordings have long been used to assist investigators in determining what happened in an accident. Speech analysis, however, holds promise for gaining insight into why it happened. The authors hope that speech analysis techniques will lead to a better understanding of cognitive and emotional states that underlie the behavior of people involved in accidents. This paper describes an initial attempt to develop a protocol for such an analysis.

Speech Measures

Speech analysis holds promise as a technique for detecting changes that may be associated with fatigue, hypoxia, alcohol intoxication, drug impairment, physical exertion, workload demand, emotional stress, and fear (Belan, 1994; Brenner & Cash, 1991; Brenner, Shipp, Doherty & Morrissey, 1985; National Highway Traffic Safety Administration, 1989). The present work is primarily concerned with the detection of workload demand and emotional stress. Several researchers have reported success in using fundamental frequency (pitch) as a measure of stress (Ruiz, Legros, & Guell, 1990; Scherer, 1981; Streeter, McDonald, Apple, Krauss, & Galotti, 1983). Brenner, Doherty, and Shipp (1994) asked subjects to count aloud while performing a tracking task with different levels of workload demand. They found that fundamental frequency and vocal intensity (loudness) increased significantly with workload demands, and speaking rate also showed a marginal increase. These measures, along with a derived measure similar to one employed by Brenner et al. (1994) and a syllable count suggested by Belan (1994), were used to analyze a speech sample from a helicopter accident. It is hoped that this work will lead to a standard protocol for speech analysis associated with accident investigation.

METHODS

The Speech Sample

On January 28, 1980, a U.S. Marine Corps UH-1N helicopter was enroute to Redding, California, on a visual flight rules (VFR) flight plan. The captain contacted a civilian Flight Service Station (FSS) by radio to exchange routine flight information and to change his destination to Red Bluff, California. Within moments of concluding this exchange, the aircraft sustained a catastrophic engine-to-transmission drive shaft failure and began an uncontrolled descent. Evidence indicated that the transmission and main rotor blades departed the aircraft during its inverted descent. The captain declared a "mayday" to the FSS and gave an assessment of the situation and a position report. The helicopter crashed shortly thereafter killing all onboard. All radio transmissions between the captain and the FSS were tape recorded by equipment at the FSS. An analysis of this recording was performed in the CVR laboratory of the National Transportation Safety Board (NTSB). (Because it involved a military aircraft, the NTSB did not conduct its own investigation of this accident.)

Analysis Procedure

The tape recording was digitized for computer-assisted acoustic analysis using an HP9000 workstation running the Waves analysis package developed by Entropic Software. Using expert guidance (Belan, 1994), statements were defined as utterances bounded by pauses of at least 300 msec. Using this definition, the sample contained 9 statements made during routine flight, and 14 statements made during the emergency. The routine statements were spoken over 46 seconds, and the emergency statements were spoken over 38 seconds; 21 seconds separated the 2 statements. Three sub-statements or phrases were spoken under both routine and emergency conditions. Five primary speech measures were made for each statement and repeated phrase: mean fundamental frequency (f_0), fundamental frequency range (Wf_0), duration, and mean amplitude (loudness) were determined with computer assistance, and the second author determined the number of syllables by listening to the digitized sample. Speaking rate (syllables per second) and 2 other derived measures were computed later. Speaking rate was not computed for utterances of fewer than 4 syllables. Following Brenner, Doherty, and Shipp (1994), the first derived measure (D-1) was computed by summing the *z*-scores of the f_0, and speaking rate for each statement. After Belan (1994), the second derived measure (D-2) was computed by summing the *z*-scores of the Wf_0, speaking rate, and syllable count for each statement (syllable counts were reverse-scored because, unlike other measures, they were expected to decrease during stress). Three analyses were conducted using these measures: (1) a statement analysis that compared f_0 and Wf_0 for each statement, (2) a condition analysis that compared routine statements to emergency statements, and (3) a phrase pair analysis that compared the phrases that were repeated under both routine and emergency conditions. (Because the radio equipment from which the recording was made was governed by an automatic gain control system, the amplitude measures were unusable in these analyses and they are not discussed further.)

RESULTS

Statement analysis (Figure 1) presents the f_0 and range of Wf_0 for each statement. The square plot symbols indicate the f_0 for each of the statements. Hollow squares indicate the 9 routine statements; filled squares depict the 11 statements made under emergency conditions. Error bars plot the range of fundamental frequencies for each statement.

It is clear from Figure 1 that the captain's fundamental speaking frequency was elevated during the emergency compared to his speech under routine conditions. Further the growth of range under emergency conditions is striking.

Condition Analysis

During routine flight, the captain's fundamental frequency averaged 123.9 Hz. This increased to an average of 200.1 Hz during emergency conditions. His Wf_0 changed from 124.2 Hz during routine flight to 297.3 during the emergency. Both of these elevations were significant using 2-tailed *t*-tests, which were used to avoid bias despite predicted difference directions. The captain averaged 11.7 syllables per statement during routine flight, but this dropped to an average of 6.7 syllables per statement during the emergency. (Six of the captain's emergency statements contained only the 2 syllable word "mayday.") If these statements are excluded, the average for the 4 remaining emergency statements is 7.8 syllables per statement. Both derived measures increased under emergency conditions, but only D-2, the Russian-influenced measure, changed significantly. Two-tailed t-tests were performed on all of these observed differences, and the results are summarized in Table 1.

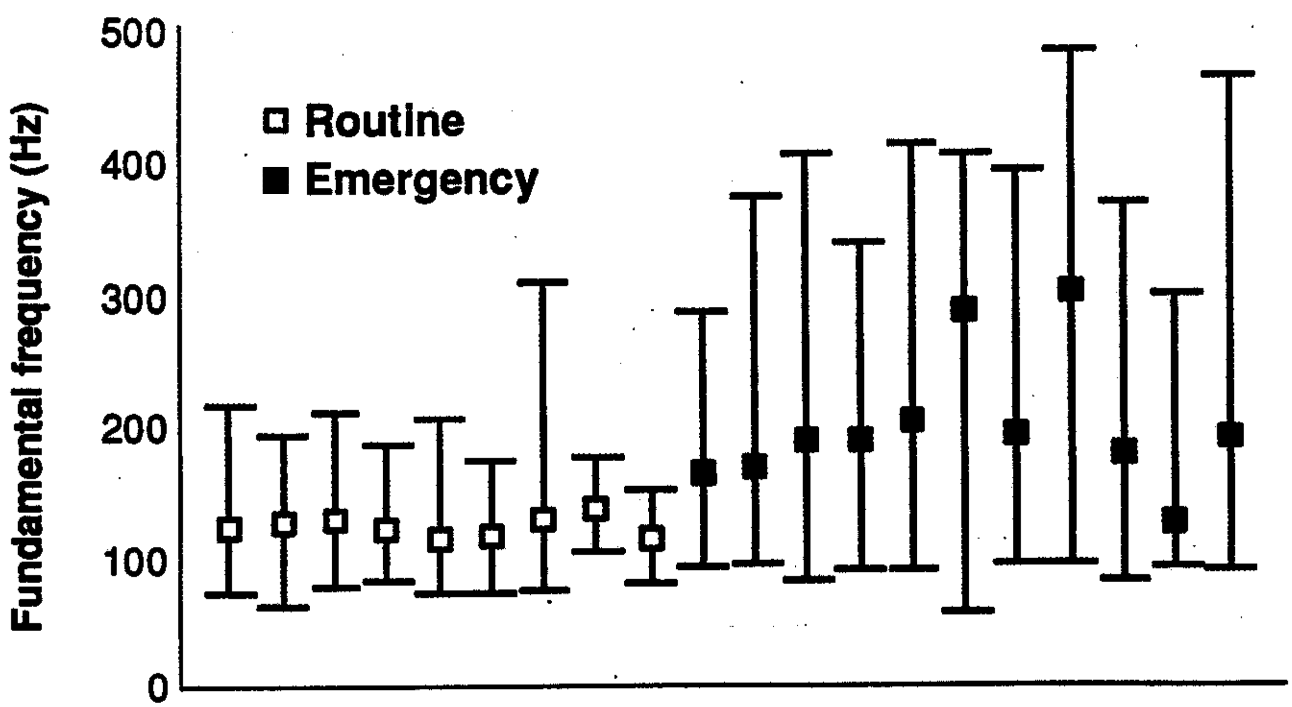

FIGURE 1: Fundamental Frequency Means and Ranges for All Routine and Emergency Statements

Measure	*Routine*	*Emergency*	*Significance*
Fundamental frequency (Hz)	**123.9**	**200.1**	**<.001**
Range of fundamental frequencies (Hz)	**124.2**	**297.3**	**<.0001**
No. of syllables	**11.7**	**6.7**	**=.054**
Speaking rate (syllables per second)	**5.3**	**4.4**	**n.s.**
Derived measure D-1	**-0.14**	**0.18**	**n.s.**
Derived measure D-2	**-0.70**	**0.90**	**<.0001**

TABLE 1: Summary of Mean Speech Measures by Condition

The information in Table 1 shows that, as predicted, both f_0 and Wf_0 increased significantly during the emergency. Also as predicted, the number of syllables per statement decreased, but this difference was not statistically significant. The derived measure used in previous work (D-1) did not change significantly, but D-2 changed dramatically. In Figure 1, the *z*-scores of the observed differences have been graphed for easy comparison. Graphical presentation of captain's speech before and during the emergency.

Phrase Pair Analysis

During the uncontrolled descent, the captain repeated 3 phrases that he had used moments earlier during routine flight. He reestablished communication by calling the FSS by its identifier, identified himself with his callsign, and gave his position. Table 2 presents speech measures for each of these phrase pairs. Although little change occurred in phrase speaking rate, large changes were seen in fundamental frequency.

Figure 2 shows the differences between the fundamental frequencies of each of phrase pairs. Each bar in Figure 2 shows the value of the fundamental frequency of one phrase, with one exception: The pilot gave his callsign twice during routine conditions; therefore, the bar that indicates this phrase actually plots the mean fundamental frequency of both phrases. A line that indicates the average fundamental frequency of all statements made during routine flight is labeled R, and a corresponding line that shows the average for all statements during the emergency is marked E (these lines plot the averages given in Table 1 for fundamental frequencies). For each phrase, the pilot's speaking pitch was higher during emergency conditions.

CONCLUSION

The extreme emotional stress experienced by the speaker during the uncontrolled descent of his aircraft is apparent in an affective sense to anyone who listens to the recording. This sample was chosen for this preliminary work because it captured 2 dramatically different emotional states, and because of the special analysis opportunities afforded by the repeated phrase pairs. The short period of time between the routine and emergency statements, and the fact that the entire recording was made using the same equipment, further made the sample attractive for this work. For these reasons, it presented a best-case scenario for development of an analysis protocol. Simply put, if the techniques described in this paper failed to work here, they would surely not work for subtler cases. The elevation in fundamental speaking frequency observed during emergency conditions is consistent with the presence of emotional stress and an increased workload demand as documented in previous studies. Further, Belan (1994) estimates that 90% of the population exhibits such a change during periods of stress.

Measure	*Routine*	*Emergency*
FSS identifier		
Fundamental freq. (Hz)	**127.3**	**193.4**
Speaking rate (syllables/sec)	**5.37**	**5.13**
Callsign		
Fundamental freq. (Hz)	**136.1**	**159.1**
Speaking rate (syllables/sec)	**6.07**	**5.38**
Position report		
Fundamental freq. (Hz)	**121.3**	**222.3**
Speaking rate (syllables/sec)	**3.21**	**4.39**

TABLE 2: Summary of Mean Fundamental Frequencies and Speaking rates for Phrase Pairs Spoken during Routine and Emergency Conditions

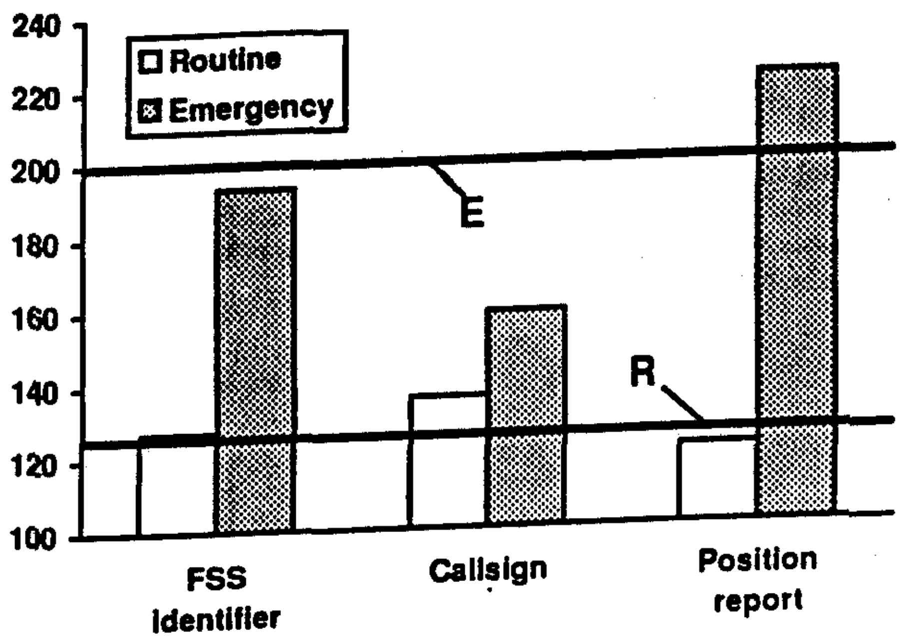

FIGURE 2: Fundamental Frequencies of Phrase Pairs, and Mean Fundamental Frequencies of all Routine (R) and Emergency (E) Statements

Further, as Belan predicted, the range of fundamental frequencies within statements grew larger under emergency conditions, and the number of syllables per statement decreased. The real value of this technique will lie in its ability to determine information about the emotional state of a speaker when it is not otherwise apparent. It is hoped that the technique described in this paper will lead to the ability to do just that in a standardized way. A tool for exploring the cognitive and emotional states of people involved in accidents could prove invaluable in determining the underlying causes of their performance and identifying appropriate preventative strategies.

Author's Note: A version of this paper was published in the Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting in October 1994. All opinions expressed in this paper are those of the authors and do not necessarily reflect the official position of the National Transportation Safety Board.

REFERENCES

Belan, A.S. (1994). Personal communication between the authors and Alfred S. Belan MD, Chief of the Acoustic Laboratory, Interstate Aviation Committee, Moscow, during his visit to the U.S. National Transportation Safety Board, Washington, D.C., February 14-18, 1994. [Dr. Belan has published his work in Russian, but it is not currently available in English.]

Brenner, M., and Cash, J.R. (1991). Speech analysis as an index of alcohol intoxication: The EXXON VALDEZ accident. *Aviation, Space, and Environmental Medicine*, 62, 893-898.

Brenner, M., Doherty, E.T., & Shipp, T. (1994). Speech measures indicating workload demand. *Aviation, Space, and Environmental Medicine*, 65, 21-26.

Brenner, M., Shipp, T.E., Doherty, T., & Morrissey, P. (1985). Voice measures of psychological stress: Laboratory and field data. In I.R. Titze and R.C. Scherer (Eds.), *Vocal fold physiology, biomechanics, acoustics, and phonatory control* (pp. 239-248). Denver: The Denver Center for the Performing Arts.

National Highway Traffic Safety Administration. (1989). Briefing paper: *Drug evaluation and classification program.* Washington, DC: U.S. Department of Transportation.

National Transportation Safety Board. (1990). *Marine accident report: Grounding of the U.S. Tankship EXXON VALDEZ on Bligh Reef, Prince William Sound, near Valdez, Alaska.* March 24, 1989 (NTSB Report No. NTSB/MAR-90/04). Washington, DC: National Transportation Safety Board. (NTIS No. PB90-916405).

Ruiz, R., Legros, C., & Guell, A. (1990). Voice analysis to predict the psychological or physical state of a speaker. *Aviation, Space and Environmental Medicine*, 61, 266-271.

Scherer, K.R. (1981). Vocal indicators of stress. In J.K. Darby (Ed.) *Speech evaluation in psychiatry*. New York: Grune and Stratton.

Streeter, L.A., McDonald, N.H., Apple, W., Krauss, R.M., & Galotti, K.M. (1983). Acoustic and perceptual indicators of emotional stress. *Journal of the Acoustical Society of America*, 73, 1354-1360.