Voice Stress Analysis

L.J.M. Rothkrantz P. Wiggers J.W.A. van Wees
R.J. van Vark

Data and Knowledge Systems Group
Delft University of Technology
Mekelweg 4, 2628 CD Delft, The Netherlands

Abstract

The non-verbal content of speech carries information about the physiological and psychological condition of the speaker. Psychological stress is a pathological element of this condition, of which one of the causes is accepted to be workload. Objective, quantifiable correlates of stress are searched for by means of measuring the acoustic modifications of the voice brought about by workload. Different voice features from the speech signal to be influenced by stress are: loudness, fundamental frequency, jitter, zero-crossing rate, speech rate and high-energy frequency ratio. To examine the effect of workload on speech production an experiment was designed. 108 native speakers of Dutch were recruited to participate in a stress test (Stroop test). The experiment and the analysis of the test results will be reported in this paper.

1 Introduction

Although speech is a vocal activity of which much is verbal, there are a number of human vocalizations that are essentially non-linguistic. Non-verbal aspects of speech are voice quality, prosody, rhythm and pausing. These phenomena stand for a non-verbal signaling system, which intertwines with the verbal or linguistic system. The non-verbal content of the voice carries, among other things, information about the physiological and psychological state of the speaker. Human beings are able to identify different emotional states, because these are characterized by clearly perceptible (non-verbal) behavior. Part of this non-verbal communication takes place via other modalities like body movements and facial expressions [9]. The question that remains is how much of this information can be recovered from non-verbal vocalizations only.

One of the most interesting research areas concerning non-verbal communication in relation to a person's psychological state is the search for objective, quantifiable correlates of stress. In the past this search focused primarily on physiological measures, but over the last years a broader range of behaviors has been examined, especially non-verbal behavior. The advantage being that stress indexes from non-verbal vocalizations can be obtained non-intrusively. From a practical point of view this is critical in a situation in which co-operation for physiological measurement is precluded, for example in the case of negotiating with terrorists.

However, even when co-operation is possible, the presence of monitoring devices needed for physiological measurement can be stressful and anxiety arousing or simply not practical.

Objective, quantifiable correlates of stress are searched for by means of measuring the acoustic modifications of the voice brought about by workload [17]. These changes in the acoustic speech signal due to stress are mainly caused by the physiological changes that accompany the stress-reaction. These changes also affect the organs of speech, such as the respiration and muscle tension (vocal cords) and therefore the speech signal. Hence, it should be possible to establish whether a person is stressed just by analyzing his voice.

2 Related Work

Much work on stress analysis in real life situations concentrates on air-ground communication in aviation and space flight under dangerous conditions. In many of these studies [1] an increase of the fundamental frequency (F0) of the voice in situations of increasing danger is reported. Williams and Stevens [1] also reported an increase in F0 range and abrupt fluctuations of F0 contour, with increasing stress. In a Russian study [12] the voices of astronauts are examined and changes in spectral energy distribution (spectral centroid moving to higher frequency) are reported. Increase of the energy of high frequency components, has also been reported by [13] in a study involving pilot communication. Scherer et al. found depressive patients speak with higher F0 and a larger proportion of high frequency components, just before the admission at a psychiatric hospital [6]. Jones [5] found increases in fundamental frequency and statistically significant decreases of the vocal jitter in recordings obtained from pilots training in a simulated AWACS environment.

In many laboratory studies, stress is brought about by showing unpleasant or disgusting slides or films, or by placing the subject in situations that produce unpleasant emotions, such as stage fright. The degree of stress perceived will vary from person to person depending on the persons experience and arousability. Apart from these individual differences, some studies show an increase in intensity, increased fundamental frequency [2, 14], stronger concentration of energy above 500 Hz [14] and an increase in speech rate [15].

More recently, many experiments were conducted in which cognitive or achievement tasks were used to induce stress on a subject [10, 7]. When persons were subjected to a psychomotor task [3], the speaking fundamental frequency showed an increase when the task became more difficult. In addition, word duration increased during the task, but decreased again when the task became more complex. Brenner [8] also found an increase in average amplitude when subjects where performing a tracking task.

Table 1 summarizes the parameters that have been shown to be indicators of the vocal expression of emotion, emotional disturbance or stress.

Table 1: Overview of major acoustic parameters [6]

Parameter	Description
F0 mean	Fundamental frequency (vibration of the vocal folds as averaged over a speech utterance)
F0 range	Difference between highest and lowest F0 in an utterance
F0 variability	Measure of dispersion of F0
F0 perturbation or jitter	Slight variations in the duration of glottal cycles
F0 contour	Fundamental frequency values plotted over time
F1 mean	Frequency of first formant averaged over an utterance
F2 mean	Mean frequency of the second formant
Intensity mean	Energy values for a speech sound wave averaged over an utterance
Intensity range	Difference between highest and lowest intensity value in an utterance
Intensity variability	Measure of dispersion of the intensity values
High frequency energy	Relative proportion of energy in the upper region
Speech rate	Length of an utterance
Spectral noise	A-periodic energy components in the spectrum
Zero crossings	Number of times a sound wave graph crosses zero

3 Experimental Design

To study the correspondence between human stress levels and speech production and to assess the relevance of the features listed in Table 1, an explorative experiment has been conducted. 108 native speakers of Dutch were subjected to several tasks that have been designed to place a cognitive workload on the subject. Cognitive workload is defined as the information-processing load placed on the human operator while performing a particular task [17]. This information processing load is considered to be correlated with the amount of attention that must be directed to a task. It is assumed that cognitive workload increases with the difficulty of the task. In the present investigation subjects performed three tasks. In the first test subjects had to play a computer game that gradually became more difficult. The second task required to simultaneously engage in two attention-demanding activities. Finally, the participants were subjected to a psychological stress test. During all tasks and during a controlled rest-condition before the tasks, the subjects produced utterances. Acoustical analyses of all utterances were made and compared with the control condition and with the acoustical analyses of the other utterances produced during the same task.

The psychological stress test, an instance of the Stroop test, proved to be the most demanding task for the subjects thus providing the clearest results. Therefore we will concentrate on the results of this task for the remainder of the paper.

3.1 Stroop Test

The Stroop test is a well-known psychological test [16] that exploits the fact that for experienced readers, the reading of a word has become an automatism. In its native form this test consisted of three cards: on the first card a great number of little squares are drawn in the colors red, blue, green and yellow. On the second card the words red, blue, green and yellow in black ink are placed on the corresponding positions. On the third card, the conflict card, the same words as on the second card are placed, but now using a non-corresponding ink-color. It turns out that the time needed to name the colors on the conflict card is much higher than the time taken for naming them on the first card. Furthermore, the subjects tend to make more mistakes reading the third card and show signs of tension (movement, sudden laughs).

In the current experiment a variation on the Stroop-test was used, in which a gradual increase of the level of difficulty is incorporated. The names of the colors (printed in different colored ink) were put on a computer screen one by one. The difficulty of the task increased as the time between the appearances of the colors was shortened every minute with half a second, thus decreasing from two and a half seconds at the start to half a second in the final minute.

3.2 Jitter

During the experiments fundamental frequency, variation of fundamental frequency, jitter, energy, high frequency energy ratio, duration and the number of zero crossings were monitored as candidate vocal stress correlates.

Jitter is the perturbation in the vibration of the vocal chords. This results in a cycle-to-cycle variation of the fundamental frequency. [4] reported that about 20 cycles are enough for jitter analysis. Formally the term perturbation implies a deviation from steadiness or regularity [11]. Let a_i be any cyclic parameter (amplitude, pitch period, etc.) in the i^{th} cycle of the waveform. Then the steady value of this parameter over a span of N cycles can be estimated from its arithmetic mean:

$$\overline{a} = \frac{1}{N} \sum_{i=1}^{N} a_i \tag{1}$$

And the zeroth-order perturbation function as the arithmetic difference:

$$p_i^0 = a_i - \overline{a}, \qquad i = 1, \ldots, N \tag{2}$$

Where the superscript gives the order of the perturbation function. Higher-order perturbation functions can be obtained by alternately taking backward and forward differences of lower order functions. We will consider the first-order perturbation function:

$$p_i^1 = p_i^0 - p_{i-1}^0 = a_i - a_{i-1}, \qquad i = 1, \ldots, N \tag{3}$$

The first order perturbation function can be used to determine the fundamental frequency perturbation if in Equation 3 a_i is taken to be the fundamental

frequency. The fundamental frequency is computed only for the voiced parts of speech. The fundamental frequency perturbation is defined as the average of the absolute values of all these differences normalised to percentage:

$$jitter = \frac{100}{(N-1)\overline{a}} \sum_{i=2}^{N} |a_i - a_{i-1}| \tag{4}$$

4 Experimental Results

In this section the results of statistical analysis of the acoustical data are described. Table 2 reports the averages and standard deviations of the data collected during the Stroop test. The conditional effects in relation to the observed effects in the features will now be discussed from condition to condition. The first minute is considered to represent normal conditions and is used for comparison.

Table 2: Results of the Stroop Test

Feature	Time	Mean	Std. dev.	Feature	Time	Mean	Std. dev.
Duration	1	454.74	136.16	High	1	52.85	15.39
	2	438.05	106.89	Frequency	2	45.06	18.50
	3	437.77	101.46		3	49.97	14.44
	4	487.66	199.72		4	32.26	17.48
	5	475.00	134.17		5	40.30	19.07
Energy	1	922.18	1386.65	Jitter	1	1.24	0.51
	2	1012.19	1307.81		2	1.05	0.49
	3	1202.56	1651.32		3	1.05	0.55
	4	884.97	1192.68		4	1.04	0.49
	5	998.17	1584.43		5	0.94	0.43
F0	1	114.28	25.26	Zero	1	3494.79	1471.62
	2	119.52	26.98	Crossings	2	3535.42	1530.14
	3	115.70	24.19		3	3522.53	1324.35
	4	122.20	27.63		4	3716.09	1404.14
	5	119.86	29.22		5	3445.05	1390.16
F0	1	7.36	12.57				
variance	2	7.38	14.12				
	3	8.59	16.42				
	4	9.86	14.78				
	5	10.11	24.05				

Stroop 2. During the second minute of the Stroop test an increase in the fundamental frequency and a decrease of the duration and jitter can be observed, as can

be expected. However, the high frequency energy shows a decrease, which is even more surprising when regarding the increase in the fundamental frequency and zero crossings. Fundamental frequency variation stays approximately the same.

Stroop 3. The fundamental frequency shows a decrease compared to the previous condition, but is still slightly higher than the first minute. Duration, zero crossing and jitter are stable at this point, but the high frequency energy ratio is still low. Fundamental frequency variation shows an increase.

Stroop 4. A steep increase in fundamental frequency and F0 variation and zero crossings is observed and still a stable jitter ratio, but high frequency is very low here. Also an increase in the duration can be witnessed, which may be because the color names differ significantly in the fourth minute of the test.

Stroop 5. A significant decrease in jitter ratio is observed in the last and most intense minute of the test. Fundamental frequency and F0 variation are still significantly higher than at the beginning of the test and duration and high frequency energy are still oppositely signed from the expected differences.

In summary the jitter ratio and partially the fundamental frequency show expected results and especially the high frequency energy shows the total opposite of what is presumed, showing an overall decrease where an increase is expected. Fundamental frequency variation shows a consistent increase towards the end of the test. However, before any conclusions are drawn the effect of the color names on the different conditions will be examined.

In Table 3 the mean values are repeated but now they are split per color. Several things can be noticed from these tables. The color *yellow* shows an overall decrease in duration (dur.) even when other colors show an increase. In all cases the jitter (jit.) shows a decrease. In practically all cases the fundamental frequency shows an increase, but differences in increase vary among the colors, which is less the case with jitter. *Blue* is the only color, which shows an increase in high frequency energy ratio. HF differences of the same colors are all closely together, which may point to dependence of the high frequency ratio on the verbal content, for example through intonational patterns. However, when looking separately at the HF's per color, there appears a reasonable consistent increase in HF over the conditions toward the end.

It can be concluded that the high frequency energy is highly dependent on the name of the color, and the jitter the least dependent. The duration is highly variable, but does show some consistency for the word *yellow*.

5 Conclusions

In this work the influence of stress on the human voice has been investigated. Stress is thought of as being caused by the workload a human is confronted with and involves a series of physiological and psychological changes. The physiological changes will among others affect speech production organs and thus the voice.

Table 3: Scores of Stroop conditions 2,3,4 and 5

Color	Nr.	Dur.	F0	Jit.	HF	Nr.	Dur.	F0	Jit.	HF
	Stroop 2					Stroop 3				
Blue	58	-10.00	2.37	-0.13	0.42	56	5.18	-1.72	-0.23	1.22
Red	4	-85.00	1.75	-0.49	-16.01	22	-51.36	6.63	-0.21	-11.48
Green	0					1	370.00	8.00	-0.13	-7.28
Yellow	24	-11.25	5.45	-0.11	-22.80	0				
Brown	0					4	45.00	5.00	-4.47	-0.13
Misses	2	-20.00	4.00	-0.79	-12.27	5	-25.00	-9.00	0.72	-12.24
Total	88	-15.68	4.41	-0.18	-7.79	88	-18.30	0.62	-0.20	-2.92
	Stroop 4					Stroop 5				
Blue	4	235.00	9.50	-0.22	-7.88	16	47.50	0.60	-0.35	3.24
Red	0					1	80.00	14.00	-1.00	-3.23
Green	21	25.71	12.53	-0.05	-34.78	12	1.67	10.73	-0.28	-37.26
Yellow	41	-8.05	4.49	-0.29	-20.66	21	36.50	2.11	-0.26	-16.54
Brown	15	24.49	2.46	-0.05	-5.85	32	10.33	2.59	-0.23	-6.10
Misses	7	241.43	7.67	-0.29	-21.39	6	238.33	10.50	-0.39	-28.46
Total	88	31.60	6.84	-0.19	-20.63	88	17.93	5.08	-0.32	12.37

Whether or not this physiological reaction pattern is to some extent person specific is still a point of discussion. It seems that the impact of a stressor is determined by his experience and physique, but this only implies that some persons can become more stressed than others.

Shifting towards the search for objective quantifiable vocal stress correlates, it turns out that a number of non-verbal vocal characteristics are subject to change when a person is speaking in a stressful situation. Among these are fundamental frequencies, duration, intensity, jitter, high frequency energy and formant positions. The experiments discussed here have shown that the most important and promising stress correlates are the fundamental frequency and the fundamental frequency perturbation or jitter. The latter is especially useful as it is relatively insensitive to prosodic patterns already present is speech, thus allowing assessment of stress levels without knowledge of the words that are being spoken. As the ultimate goal of our research is to develop a stress-o-meter based on non-intrusive techniques such as speech recognition we are planning to assess how well the vocal stress features correlate with physiological measurements [18].

References

[1] K.N. Stevens C.E. Williams. On determining the emotional state of pilots during flight, an exploratory study. *Aerospace Medicine*, 1969.

[2] L.A. Streeter et. al. Pitch change during attempted deception. *Journal of Personality and Social Psychology*, 1977.

[3] C.E. Williams G.R. Griffin. The effects of different levels of task complexity on three vocal measures. *Aviation, Space and Environmental Medicine*, 58:1165–70, 1987.

[4] H. Liang I.R. Titze. Comparison of f_0 extraction methods for high-precision voice pertubation measurements. *Journal of Speech and Hearing Research*, 36, 1993.

[5] W.A. Jones. *An evaluation of voice stress analysis techniques in a simulated AWACS environment.*

[6] F. Tolkmitt K.R. Scherer. The effect of stress and task variation on formant location. *Journal of the Acoustical Society of America*, 1979.

[7] G.E. Schwartz M. Brenner, H.H. Branscomb. Psychological stress evaluator: Two tests of a vocal measure. *Psychophysiology*, 1979.

[8] T. Shipp M. Brenner. Voice stress analysis: Mental state estimation. In J.R. Construck, editor, *NASA conf. pub. 2504*, 1987.

[9] L.J.M. Rothkrantz M. Pantic. Towards an affect-sensitive multimodal human-computer interaction. *Proc. of the IEEE*, 91(9):1370–1390, 2003.

[10] C.E. Williams M.H.L. Hecker, G. von Bismarck. Manifestations of task induced stress in the acoustical speech signal. *Journal of the Acoustical Society of America*, 44:993–1001, 1968.

[11] I.R. Titze N.B. Pinto. Unification of pertubation measures in speech signals. *Journal of the Acoustical Society of America*, 87:1278–89, 1990.

[12] M.V. Frolov P.V. Simonov. Utilization of human voice for estimation of man's emotional stress and state of attention. *Aerospace Medicine*, 1973.

[13] J.W. Lester R. Roesller. *Vocal patterns in anxiety.* 1979.

[14] K.R. Scherer. The effects of stress on the fundamental frequency of the voice. *Journal of the Acoustical Society of America*, 1977.

[15] A.W. Siegmann. *Paraverbal and non-verbal indicators of stress.* 1992.

[16] J.R. Stroop. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 28:643–662, 1935.

[17] H.G. Strassen T.B. Sheridan. *Definitions, models and measures of human workload.* Plenum, New York, 1979.

[18] R.J. van Vark. *Knowledge based behaviour feedback system using physiological data, master's thesis.* TU Delft, 1993.