VOCAL JITTER AS AN INDICATOR OF CHANGES IN PSYCHOPHYSIOLOGICAL AROUSAL

Erkki Vilkman **Olavi Manninen** **Eija-Riitta Lauri** **Tarmo Pukkila**

Phoniatric Dept.,
Tampere University
Central Hospital,
SF-33520 TAMPERE
FINLAND

Dept. of Public Health,
University of Tampere

Phoniatric Dept.,
Helsinki University
Central Hospital

Dept.of Mathematical
Sciences,
University of Tampere

ABSTRACT

The possibilities of using cycle-to-cycle changes in fundamental frequency (jitter) for estimating changes in subjects' psychophysiological arousal were studied. The subjects (n=20) were exposed to four different combinations of dry bulb temperature (20°C or 35°C), noise (90 dBA) and whole-body vibration (sinusoidal 5 Hz) in a special exposure chamber. The exposure lasted the whole day. The jitter was measured manually of an excerpt from a text which the subjects read during rest and exposure separately from morning and afternoon samples. Only in the afternoon samples did the changes in jitter caused by exposure to 20°C (increase) and 35°C (decrease) temperatures differ significantly (p<0.05, df=8).

INTRODUCTION

The tremor of one's voice is a well-known feature of anxiety, prompted, for example, by performing situations. Thus it is natural that there have been attempts to use voice changes, for instance, in detecting deceptive behavior. Commercially available apparati which have been claimed to reflect a person's emotional state have been developed in this field. The most popular of these is the so-called Psychological Stress Evaluator (PSE), which in addition to forensic science /1,2/ has also been used in the field of psychology /3/.

In brief, PSE consists of an electrical integration circuit which filters the acoustic signal so that a fluctuation at a rate of about 10 Hz of the baseline can be seen. In anxiety states this tremor is claimed to diminish /1,2/. The origin of the tremor is somewhat obscure at present. The reliability and validity of PSE has been seriously questioned /1,2,4/.

Another fluctuation phenomenon of the human voice is the cycle-to-cycle variation in the fundamental frequency, i.e. the jitter. Jitter has been studied rather extensively as a sign of vocal pathology /5/, but it has also been found to be associated with the emotional contents of speech /6/. It has been claimed that in so-called "stress situations" the jitter scores tend to lower with increasing threat /7/.

Temperature, noise and vibration are important exposure factors for research, because in modern society people are exposed to them almost daily. However, there is relatively little knowledge concerning their individual effects and hardly any concerning their combined effects on the psychophysiological arousal of human subjects. In an earlier study on changes in prosodic features of speech due to environmental factors /8/, we noticed that different combinations of temperature, noise and whole-body vibration caused changes in the average fundamental frequency, intensity, spectral characteristics and durational variables of speech.

The changes in the prosodic features could be interpreted in terms of existing knowledge of psychophysiological changes related to similar exposure conditions /8/. The aim of this preliminary study was to determine whether jitter could be used as a sensitive indicator of psychophysiological arousal. For this purpose we analysed excerpts of speech samples from four different exposure combination cells of the earlier study.

SUBJECTS AND METHODS

Excerpts of reading samples of twenty healthy male subjects were analysed. The subjects (n=20; 5 in each) were exposed to the following exposure combinations: 1) 20°C temperature (T), no noise (N), no vibration (V) (T1N0V0); 2) 35°C temperature only (T2N0V0); 3) 35°C and 90 dB(A) noise (T2N1V0) and 4) 35°C temperature, 90 dB(A) noise and 5 Hz sinusoidal whole-body vibration along Z-axis (T2N1V1). The experiment was carried out in a special exposure chamber. During the test, subjects sat in a vibration chair. See /9/ for a detailed description of the exposure arrangements.

The samples of speech were recorded during a pause in the test. Altogether four speech samples were recorded. Rest session samples were recorded from 9:13 to 9:15 in the morning before the exposure and in the afternoon from 12:13 to 12:15. The first exposure sample was recorded in the morning between 11:08 and 11:10 after an exposure session lasting 80 minutes. The second was recorded in the afternoon at 2:08-2:10 after total exposure (lasting altogether176 minutes). See /8,10/ for the recording arrangements and further details.

The excerpt on which the cycle-to-cycle analysis was performed consisted of vowels and voiced consonants (/on lauan.../). This excerpt starts a new chapter and is very emphatic. The excerpt ends at a voiceless consonant(/t/). The sampling rate of the signal was 10 kHz. In the analysis of recorded sinusoidal sounds the jitter of the apparati fell below the accuracy of the measurements. The measurements were carried out blindly. On an average (X±SD) 45.7±9,8 successive periods were measured. Measurements were made manually from oscillographic displays by means of a cursor. The results of the measurements were stored on disk and drawn on plotting paper using microcomputer-based (Motorola Exorset) programs.

The jitter value was formed as the difference between the observations and the five point moving average to avoid the influence of the general trend of the fundamental frequency. If the original observations are denoted by X_t, the five-point moving average value is obtained from the formula

$$T_t = (X_{t-2} + X_{t-1} + X_t + X_{t+1} + X_{t+2})/5.$$

The jitter J_t is therefore obtained as the difference

$$J_t = X_t - T_t$$
$$= (- X_{t-2} - X_{t-1} + 4X_t - X_{t+1} - X_{t+2})/5,$$

i.e. J_t is obtained as a five-point moving average such that the sum of the weights is zero.

The standard deviation (SD) of the jitter J_t was used as the jitter index [J(*tot*), J(*asc*)]. The mean of the fundamental frequency (Fj) of the measured excerpt was calculated. The analyses were run separately for the ascending part (ΔJ(*asc*) (pitch rise) and the total excerpt(ΔJ(*tot*) (c.f. Fig.1).

Before statistical treatment, both the morning and afternoon exposure sample results of jitter were standardized by subtracting the results of the preceding rest samples. The correlation of jitter [ΔJ(*tot*)] with the changes in fundamental frequency (ΔF0), intensity (ΔI) total reading time (ΔTt), articulation time (Ta), and number and errors in /tapaka/ and /pataka/ word repetitions obtained from the earlier study /8/ was calculated.

RESULTS

The mean and the standard deviation of Fj of the excerpt was 112.5±19.3 Hz The mean J(*tot*) value of the total excerpt was 2.9 Hz and the mean J(*asc*) of the ascending part 2.0 Hz.

Fig. 1. shows an example of FO curves in rest and exposure.

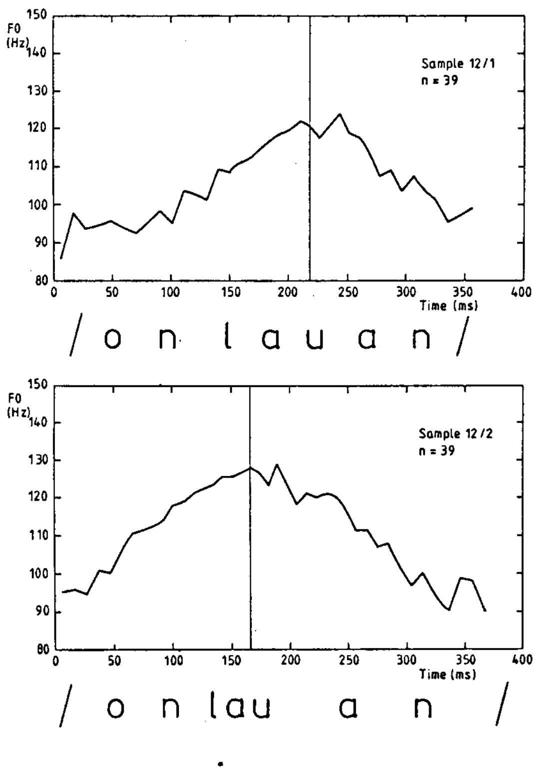

Fig. 1. The F0 curves measured cycle-to-cycle on the excerpt from the text (/on lauan.../) for one subject (exposure combination T2N1V0). The jitter is bigger for rest (top) [J(*tot*)=2.29 Hz, J(*asc*)=2.39 Hz] than exposure (bottom) [J(*tot*)=1.54 Hz, J(*asc*)=1.21 Hz]. F0 is higher in the exposure sample. The vertical line shows the estimated end of the ascending part (start of /u/) of the curve which was analysed separately.

The mean changes in the measurements of the excerpt are shown in Table I by exposure combinations. It can be noted that the changes in jitter values [ΔJ(*asc*) and ΔJ(*tot*)] are small in general and that interindividual variation is big. Differences between the effects of the exposure in the morning and afternoon samples are also clear. In the morning samples the clearest drop in jitter [ΔJ(*tot*)] has occurred due to exposure to combination T2N1V0. This combination raised F0, I, Tt and Ta values of the total sample, and the rate of word repetitions (/tapaka/ and /pataka/) increased /8/. In the afternoon samples the effects of T1N0V0 and T2N0V0 on ΔJ(*tot*) differed statistically significantly (p<0.05, t=2.32, df=8).

<u>Table I.</u> Average changes (X±SD) in jitter and fundamental frequency measurements of the text excerpt. See text for symbols.

	ΔJ(*asc*)	ΔJ(*tot*)	ΔF(j)
a.m.	(Hz)	(Hz)	(Hz)
T1N0V0[1]	-0.5±1.4	0.5±2.0	-1.1.±2.9
T2N0V0	-0.2±0.9	-0.1±1.4	6.1±7.1
T2N1V0	-0.7±1.2	-2.0±1.9	5.3±6.5
T2N1V1	0.2±0.7	-0.6±1.5	7.8±5.9
p.m.			
T1N0V0	0.0±1.4	1.5±2.5	4.4±4.4
T2N0V0	-0.4±1.3	-1.3±1.2	5.1±3.0
T2N1V0	0.3±0.8	0.0±1.0	7.7±6.8
T2N1V1	0.2±0.7	-0.4±2.5	1.2±5.6

[1] Exposures were T1=20°C, T2=35°C temperature, N0=no noise, N1=90 dBA, V0=no vibration, V1= 5 Hz sinusoidal vibration.

Table II shows the correlation coefficients of jitter measurements with prosodic features of speech. It can be seen that in general the correlations are weak. The ΔJ(*tot*) values, however, show a tendency to negative correlation with ΔF0 and ΔI. Very interesting are the correlations between ΔJ(*tot*) and the /pataka/ word repetitions and especially the statistically significant correlation with the number of errors in the afternoon samples and consequently also in a.m.+p.m. calculations. ΔJ(*tot*) also correlated weakly with the ΔFj value: in the morning samples the correlation was negative (r=-0.26) and in the afternoon positive (r=0.32).

<u>Table II.</u> Product moment correlation coefficients (*r*) between changes in jitter and prosodic variables. See text for symbols.

					Δ'tapaka'		Δ/pataka/	
	ΔF0	ΔI	ΔT t	ΔTa	n:o	errors	n:o	errors
a.m.								
ΔJ (*tot*)	-0.28	-0.14	0.03	-0.25	-0.15	-0.11	-0.31	0.17
ΔJ (*asc*)	-0.15	0.07	-0.12	-0.22	-0.01	-0.35°	-0.19	0.18
p.m.								
ΔJ (*tot*)	-0.24	-0.28	0.05	0.08	0.17	0.05	-0.17	0.58**
ΔJ (*asc*)	-0.04	0.09	-0.14	0.16	0.02	0.33	-0.07	0.25
a.m.+p.m.								
ΔJ (*tot*)	-0.25	-0.29°	0.05	-0.03	0.02	-0.03	-0.24	0.43**
ΔJ (*asc*)	0.05	0.07	-0.10	0.02	-0.01	0.13	-0.16	0.23

° $p<0.10$, * $p<0.05$, ** $p<0.01$, a.m.&p.m. df=20, a.m.+p.m. df=40

DISCUSSION

The changes in fundamental frequency (ΔFj) of the excerpt did not show significant exposure-specific changes. This may be due to the fact that such an average value reflects more the changes in reading style, which is more complicatedly related to changes in psychophysiological arousal, than the supposed tension changes underlying the long-time average F0 (c.f. /11/).

A five-points moving average was used to calculate the jitter. According to Kitajima et al. /12/ the use of four points already produced a satisfactory smoothening effect. The absolute values of the jitter of the present study cannot be compared to other studies because of the differing types of jitter indices /13/. The jitter values of the present study may be biased due to a relatively low sampling rate, which has been found to increase the magnitude of jitter /5/ and the random effect of the jitters of the recording and the analysis system (c.f. /7/).

The magnitude of the physiological jitter has been reported to be related to the asynchronous firing of the motor units of the cricothyroid muscle; thus the decrease in jitter is caused by an increase in neural input and arousal /7,14/. A drop in jitter values /6, 7/ and a rise in F0 and intensity /4,15,16/ have been noticed to be connected with psychological stress. In terms of these reports it can be supposed that the exposure to the combination of 35°C temperature and 90 dB(A) noise caused a rise in arousal in the morning samples. The effect of T2N0V0 on the ΔJ(*tot*) in the afternoon samples also implies psychological stress.

The positive correlation between the jitter and errors in /pataka/ word repetitions might imply a common basis for these changes. In an earlier study we found that /pataka/ word repetitions and errors might be useful in assessing changes in a person's arousal /8/. It can be hypothesized that the increase in magnitude of jitter is associated with lowered arousal, and thus the tendency to increased errors can be interpreted in terms of motivation (the correlation was highest in the afternoon samples). And vice versa, the performance becomes better with higher arousal and an increase in motivation.

In conclusion, the results of this preliminary study suggest that the jitter is not a more sensitive indicator of a person's psychophysiological arousal than the prosodic features of speech. However, it has to be kept in mind that different vocal and speech variables may reflect some independent specific changes in arousal due to the exposure. This would be in line with recent psychophysiological studies on the demand-specificity of changes in various physiological measures /17/.

REFERENCES

/1/ VanDercar DH, Greaner J, Hibler NS, Spielberger CD, Bloch S. A description and analysis of the operation and validity of the psychological stress evaluator. J Forens Sci 25:174-188, 1980.
/2/ Horvath F. Detecting deception: the promise and reality of voice stress analysis. J Forens Sci 27:340-351, 1982.
/3/ Smith GA. Voice analysis for the measurement of anxiety. Br J Med Psychol 50:367-373, 1977.
/4/ Hollien H. Vocal indicators of psychological stress. pp. 47-72. In: Wright, Bahn, Rieber (eds.) Forensic psychology and psychiatry. Ann N Y Acad Sci vol 347, 1980.
/5/ Heiberger VI, Horii Y. Jitter and shimmer in sustained phonation. Speech Lang: Adv Basic Res 7:299-332, 1982.
/6/ Lieberman P. Perturbations in vocal pitch. J Acoust Soc Am 33:597-603, 1961.
/7/ Brenner M, Shipp T, Doherty T, Morrissey P. Voice measures of psychological stress. pp. 240-248. In: Titze, Scherer (eds.) Vocal fold physiology, Center for Perf Arts, Denver 1983.
/8/ Vilkman E, Manninen O. Changes in prosodic features of speech due to environmental factors. Speech Comm 5: 331-345,1986.
/9/ Manninen O. Hearing threshold and heart rate in men after repeated exposure to dynamic muscle work, sinusoidal vs stochastic whole body vibration and stable broadband noise. Int Arch Environ Health 54:19-32, 1984.
/10/ Vilkman E, Manninen O. Changes in prosodic features of speech under complex exposure conditions. pp. 145-167. In: Manninen (ed.) Combined effects of environmental factors, Proc Ist Int Conf Combined Effects of Environmental Factors, Tampere 1984. Keskuspaino, Tampere, Finland 1984.
/11/ Scherer K, Ladd DR, Silverman KEA. Vocal cues to speaker affect: testing two models. J Acoust Soc Am 76:1346-1356, 1984.
/12/ Kitajima K, Tanabe M, Isshiki N. Pitch perturbation in normal and pathological voices. Studia Phonol 9:25-32, 1975.
/13/ Zyski BJ, Bull GL, McDonald WE, Johns ME. Perturbation analysis of normal and pathological larynges. Folia Phoniat 36:190-198, 1984.
/14/ Baer T. Vocal jitter: A neuromuscular explanation. pp.19-22. In: Lawrence, Weinberg (eds.) Transact 6th Symp Care Prof Voice. The Voice Foundation, New York, 1980.
/15/ Scherer K. Vocal indicators of stress. pp. 189-220. In: Darby (ed.) Speech evaluation in psychiatry. Grune & Stratton, New York, 1981
/16/ Williams CE, Stevens KN. Vocal correlates of emotional states. pp. 221-240. In: Darby (ed.) Speech evaluation in psychiatry. Grune & Stratton, New York, 1981.
/17/ Lyytinen H. The psychophysiology of anticipation and arousal. PhD Diss. University of Jyväskylä, Jyväskylä, Finland, 1984.

ACKNOWLEDGEMENTS

This study was financed by the Academy of Finland. The authors wish to express their gratitude to Professor Jaakko Lehtonen for technical and practical support.