Proceedings of the Institute of Acoustics

EFFECTS OF WHOLE-BODY VIBRATION ON SPEECH

M Goldsmith (1) & M Taylor (2)

(1) National Physical Laboratory, Teddington TW11 0LW, UK.
(2) now with Ferranti International PLC, Bracknell, RG12 2SJ, UK.

1. INTRODUCTION

One of the most useful applications for an advanced automatic speech recogniser would be the cockpit of an aircraft, where the high workload on the pilot could be considerably reduced by offloading some peripheral tasks to the recogniser. Unfortunately, this operating environment is also one of the most hostile, combining as it does high levels of noise, vibration, acceleration and emotional tension.

Knowledge of the effect of such conditions on the human voice, and hence on recogniser performance, is essential if recognisers are to be used in such environments. The National Physical Laboratory, in collaboration with Smiths Industries Aerospace and Defence Systems, has completed a pilot study of the effect on the voice of whole body vibration. The data was collected on a vibrating test-rig at the Royal Aircraft Establishment, Farnborough. In particular, the effect of vibration on formant frequencies was investigated.

It is fairly well established that one major effect of physical and/or emotional stress on speech is an increase in pitch (Nixon [1], Williams & Stevens [2]), and one of the chief components of physical stress in aircraft is vibration, and in particular, vibration at frequencies below 25 Hz (Taylor [3]). This fact has gradually been quantified, most recently by Taylor, who has studied the effects of different vibrations induced in pilots on a three-axis vibrating rig on human speech. He found (Taylor [3]) that the vibration frequency which produced the strongest effects (in particular the largest shift in fundamental frequency) was 9 Hz. The present paper investigates in more detail the effects on the speech signal at this frequency, by statistical analysis of the larynx signal.

2. VIBRATION EXPERIMENTS

Ten adult male speakers were used in the experiments; eight had South East accents, while one was from Tyneside and the other Scots. Ages ranged from 20 to 47 (mean age was 32.5). Speaker information concerning health, habits and experience was recorded (for details see Taylor [3]): about half were smokers, a minority were pilots, about half had had their tonsils removed and none had any throat problem or infection. The speakers sat in the seat of a Chinook helicopter, which was fixed to the rig and had a resonant frequency at 27 Hz - too high to be excited by the frequencies applied here. Throughout the experiment, acoustic speech signals, laryngograph signals, and values of displacement were recorded (via a multiplexer) on a Sony Betamax standard video recorder. The multiplexer channels and bandwidths were as follows:

Channel 1: Inertial Resultant Gz. DC - 3 kHz = 3 dB (8 BITS)
Channel 2: Inertial Resultant Gx. DC - 3 kHz = 3 dB (8 BITS)
Channel 3: Inertial Resultant Gy. DC - 3 kHz = 3 dB (8 BITS)
Channel 4: Spare
Channel 5: Laryngograph Lx Data. 10 Hz - 10 kHz = 3 dB (8 BITS)
Channel 6: Speech Channel. 20 kHz (14 BITS)

The subjects were vibrated at a range of frequencies and forces (from 8 to 25 Hz, and from 0.05 to 0.25 g rms respectively), while speaking a set of cockpit command and control passages chosen to be as similar as possible to what would be expected of pilots when controlling cockpit equipment via a speech recogniser. There were three long sentences and 48 short ones, and the whole list took between two and three minutes to read and was read once for each combination of acceleration and vibration. Control recordings were also made under stationary conditions. No subject was exposed to more than 30 minutes total vibration, without a rest of at least two hours before continuing.

The larynx signals corresponding to the passages were converted to sequences of glottal pulse-durations, and statistical analyses of these datasets were undertaken: the data were sorted into 128 ranges of voice-lengths, allowing the construction of frequency histograms. The ranges were logarithmically spaced, in order to ensure that the histograms of speakers with different fundamental frequencies would occupy roughly the same number of bins. Noise led to the appearance of both narrow spikes and longer term low level spurious data in the larynx signal. The short-term irregularities were removed by rejecting data points which were more than 10% longer or shorter than previous or subsequent ones, since such rapid changes could not occur naturally in larynx signals. This process removes typically 2-10% of data. The low level contaminants of the larynx signal were removed by setting a noise level of 0.25%. From the "cleaned" histograms that were thus produced, values of mean and modal frequency, skew, range and kurtosis were determined. Kurtosis is of particular interest since it has proved to be a good indicator of fluency/monotony, while skew indicates the relative usage of available frequency ranges (see Barry et. al [4] for discussion). Values of the ratio of total time for which the voice pitch was rising to the total time during which it was falling were also calculated.

Statistics were calculated for the speakers at rest and under 9 Hz vibration at maximum force (0.25 g), and the results are compared below:

PARAMETER	0 Hz	9 Hz	Difference	Significance
Mode (Hz)	136	160	+18%	95%
Mean (Hz)	133	160	+19%	95%
Mid Range Point (Hz)	135	160	+19%	95%
Start of Range (Hz)	113	129	+14%	N.S.
End of Range (Hz)	156	191	+22%	95%
Range (octave)	.32	.39	+43%	95%
Skewness	-.21	.04	+.25	Marginal
Kurtosis	3.3	3.2	+4%	N.S.
Rise/Fall Ratio	.50	.56	+13%	N.S.

3. DISCUSSION

It is clear from the above that all measures of central tendency demonstrate a significant rise in frequency with applied vibration. In addition there is a significant increase in range, due to the use of higher than usual frequencies in the voice. There is also a marginally significant (90% level) increase in skew, suggesting that these high frequencies account for only a small part of the total amount of signal. No significant effect was found on kurtosis or rise/fall ratio.

The main cause of general voice distortion by 9 Hz vibration is resonance of the anterior chest and abdominal wall (Coermann et al [5]), while the chief cause of the increased voice pitch is likely to be an increase in muscular tension due to an unconscious effort to counteract these body resonances. The vocal fold tension would also increase, raising the larynx frequency (Nixon [1], Judson & Weaver [6]).

It is interesting to note that while such increased pitch is clearly present, use is still made of the lower frequencies, as the unchanged lower limit of frequency range shows. This factor may mean that speech recognition algorithms that exploit lower frequencies may perform better in vibrating environments than algorithms which depend on higher frequency components of the signal.

Note: while this work was in progress, M Taylor was employed by Smiths Industries Aerospace and Defence Systems, under secondment to the National Physical Laboratory.

REFERENCES

[1] C W NIXON, 'Influence of Selected Vibrations upon Speech', Journal of Auditory Research, 2 p247 (1962)
[2] C E WILLIAMS & K N STEVENS, 'Emotions and Speech: Some Acoustical Correlates', J Acoust Soc Am, 52 p1238 (1972)
[3] M R TAYLOR. 'Studies in Automatic Speech Recognition and its Application in Aerospace', Ph D thesis (1989)
[4] W J BARRY, M J GOLDSMITH, A J FOURCIN & H C FULLER, 'Larynx Analyses of Normative Reference Data', STA Project Report (1990)
[5] R R COERMANN, G H ZIERGENRUEKER, A L WITTWER & H E VON GIERKE, 'The Passive Dynamic Mechanical Properties of the Human Thorax-Abdomen System and of the Whole Body', Aerospace Medicine, 31 p443 (1960)
[6] L S JUDSON & A T WEAVER, 'Voice Science' (1942)

ISBN 1 873082 13 4 pp. 283 - 581
ISBN 1 873082 14 2 set of two