SOME CHARACTERISTICS OF SPEECH PRODUCED UNDER HIGH G-FORCE AND PRESSURE BREATHING.

Allan J. South
Systems Integration Dept.,
Defence Evaluation and Research Agency,
Farnborough, Hampshire, GU14 0LX, UK

Abstract

The performance of speech recognisers in combat aircraft is degraded seriously by the extreme physical stresses to which the crew are subjected. This paper describes measurements of first and second formant frequencies of nine vowels from one speaker recorded under high levels of acceleration, with and without positive pressure breathing. Under acceleration alone, F2 is reduced for high front vowels, while F1 remains constant, but for back and mid vowels, F1 reduces with little change in F2. When positive pressure breathing is introduced, nearly all vowels are affected, and the "vowel triangle" on the F1-F2 plane collapses inwards, towards the neutral vowel position. If these changes are found to be consistent between speakers, it is hoped to develop techniques of voice transformation to reverse them, and thus improve the performance of speech recognisers in this harsh environment.

1. INTRODUCTION

Agile combat aircraft currently under development (e.g. Eurofighter) will use speech recognition as a part of the cockpit interface. However, the extreme physical stresses placed on the pilot in such an aircraft will cause changes in his speech and thus adversely affect the performance of the speech recogniser. One way to overcome this problem may be to adapt the recogniser's word models to the current environment using information on G level and breathing gas pressure supplied by the aircraft's systems. For this to be successful, detailed knowledge of the effects of these stresses will be needed. Previous work [1,2] has given some indications of the effects of high levels of acceleration, but the number of samples examined has been limited, and, as far as is known to this author, the effects of positive pressure breathing have not been studied at all. This paper describes measurements of formant frequencies carried out on speech recorded at up to 6g in a man-carrying centrifuge, as a preliminary to developing methods of compensation for these effects.

2. SPEECH RECORDINGS

Recordings were made by five male subjects, all Royal Air Force personnel, at 1g increments up to 8g [3]. For one session, at 4g and above, positive pressure breathing was used, with a chest counter-pressure garment. Pressure breathing involves increasing the pressure of the breathing gas supply above that of cockpit; this causes the pilot's blood pressure to increase, thus allowing him to stay conscious and function under accelerations as high as 9g. The pressure schedule used during these recordings raised the pressure in the mask by 10 mmHg per g from the safety pressure level of 4 mmHg at 2g

Lists of digit strings and phrases typical of a cockpit voice control task were recorded, including 25 utterances in each case. Subjects wore a standard RAF oxygen mask, a lightweight MK10 flying helmet, anti-G trousers and life jacket. The mask microphone was connected to a standard cockpit intercom control unit, the output signal from this being recorded on Digital Audio Tape. Recordings were also made off the centrifuge for recogniser training; these recordings were made with no pressure in the mask, whereas the 1g condition recorded on the centrifuge had safety pressure of 4 mmHg in the mask.

The original purpose of these recordings was to collect data for testing the performance of automatic speech recognisers under high G conditions with syntax and vocabulary typical of a fast-jet cockpit application. The recognition performance reduced gradually as the g-level increased, but fell steeply when pressure breathing was introduced [3].

The recordings were later downsampled to 8 kHz and filtered to compensate for the frequency response of the mask and microphone [4], although the filtering did not completely eliminate the sharp peak in the microphone's response at about 2.8 kHz (see Figure 1). The compensation was necessary in order to optimise the performance of the formant tracker, and automatic labelling software. The files were labelled at word and phone level.

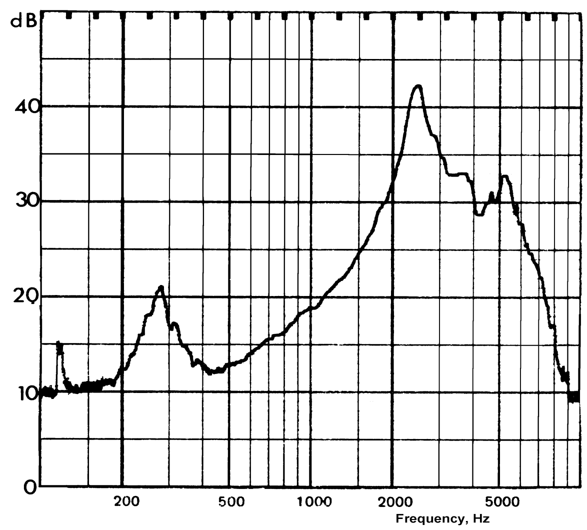

Figure 1 Typical microphone frequency response

3. FORMANT FREQUENCY MEASUREMENTS

3.1 Selection of vowels for analysis

Formant tracks were generated automatically from the recordings of one speaker, using a proprietary formant tracker. After pre-emphasis (which was not included in the compensation filter), estimates of the first four formant frequencies were derived from the roots a 14th order linear predictor polynomial, then optimised using dynamic programming to meet frequency continuity constraints. A window length of 20 ms was analysed every 10ms. All formant tracks were checked manually. First and second formant frequencies were extracted for the vowels /@, i, I, e, {, V, Q, O, u/. Since the recorded material was not designed specifically for acoustic-phonetic analysis, the number of examples of each phone is limited and variable. Ideally, one would like several examples of each vowel in the same context. Some, such as /e/ and /@/ are plentiful, as the word "enter" occurs at the end of most of the 25 phrases; others, such as /u/, occur only a few times.

An additional complication arises for vowels on word boundaries, for which the context may vary depending on the adjacent word and whether the subject pauses between the words. It has been our experience that RAF aircrew usually insert slight pauses between words, even when instructed to speak "naturally". One effect of this has been that embedded training has shown no advantage over isolated training in recognition tests on our airborne recordings [5]. It may be assumed therefore that context effects at word boundaries are much less significant in the speech being studied here than in more natural speech.

As a general rule, only vowels occurring in the list at least four times in the same context have been included in the analysis. An exception was made in the case of /O/, which occurs in the words "FOUR" and "FORMAT". Of the four instances of "FOUR" in the list, the following word is "ENTER" in three cases and "MINUTES" in one case. The values of both F1 and F2 from all instances in both of the 1g lists were subjected to a t-test, which showed no significant differences between the contexts (p>30%). Examination of the spectrograms shows that the duration of /O/ is so long that the target values are always reached, which would suggest that the context has little effect on the formant frequencies measured at the centre of the vowel.

The vowel /A:/ occurs only once in the list of command phrases and not at all in the digits, and so has been excluded from this analysis.

Table 1 summarises the contexts of the vowels studied. In all cases, the accuracy of the formant tracks was checked manually and corrected where necessary. The formant frequency measurements were generally taken in the centre of the vowel, but for /i, V, u/ the measurement was taken towards the end when the formant frequencies had usually reached stable values.

Vowel	**Context**		**No. of utterances per condition**
	Orthographic	**SAMPA**	
/@/	enter	/ent@/	10
/i/	three	/Tri/	8
/I/	MIDS	/mIds/	4
/e/	enter	/ent@/	10
/{/	format automatic	/fOm{t/ /Ot@m{tIk/	2 2
/V/	one	/wVn/	4
/Q/	box	/bQks/	5
/O/	four format	/fO/ /fOm{t/	4 2
/u/	two	/tu/	4

Table 1 Vowels used for the analysis

3.2 Conditions

The original recordings were also intended to assess the effects of different types of anti-G clothing, so up to three readings of the lists were available at each level of acceleration. There were no significant differences in the accuracy of the speech recogniser between the different types of G protection, so in order to reduce the amount of data to be analysed, one list was chosen for each level of acceleration from 1g to 5g. The introduction of positive pressure breathing had a very significant effect on the recognition accuracy, so the lists at 4g, 5g, and 6g with pressure breathing were also studied.

Analysis of third and fourth formant frequencies has not been attempted. The filter used to compensate for the microphone

Vowel	F1, Hz		F2, Hz	
	Mean	Std Dev	Mean	Std Dev
/@/	643	22	1287	47
/i/	396	53	2079	62
/I/	438	53	1909	41
/e/	640	42	1704	55
/{/	730	28	1298	84
/V/	680	8	1088	13
/Q/	599	27	854	31
/O/	480	8	759	41
/u/	435	46	1609	42

Table 2 Mean formant frequencies in Hz at 1g

	/@/		/i/		/I/		/e/		/{/		/O/		/u/	
Acc.	F1	**F2**	F1	**F2**	F1	**F2**	F1	**F2**	**F1**	F2	**F1**	F2	F1	F2
1g	643	**1287**	396	**2079**	438	**1909**	640	**1703**	**745**	1353	**480**	759	435	1609
2g	597	**1349**	374	**1988**	434	**1836**	657	**1580**	**752**	1396	**498**	747	388	1713
3g	600	**1416**	347	**1991**	400	**1791**	606	**1675**	**692**	1363	**444**	744	350	1757
4g	641	**1426**	359	**1920**	443	**1747**	590	**1605**	**720**	1315	**465**	746	376	1716
5g	645	**1481**	376	**1892**	473	**1782**	631	**1438**	**689**	1388	**436**	741	505	1442

Table 3 Mean F1 and F2 for vowels produced under acceleration with safety pressure only (units Hz)

frequency response does not completely remove the strong peak at about 2.8 kHz, with the result that the formant tracker often places one of these formants at this frequency, which is not related to the configuration of the vocal tract.

4. RESULTS

4.1 Baseline results at 1g

Table 2 shows the mean values and standard deviations of F1 and F2 for each vowel, measured from the 1g recordings i.e made with the centrifuge stationary.

4.2 Variation of F1 and F2 with acceleration

Table 3 shows the mean frequencies for the first and second formants for the five conditions recorded on the centrifuge, but without positive pressure breathing (other than safety pressure). Assuming a linear dependence of formant frequency on acceleration, regression tests were carried out on these values. Formants showing a statistically significant dependence on acceleration (H0: slope=0; $p < 5\%$) are shown in bold face. The vowels /V, Q/ showed no dependence on acceleration for either formant and have been omitted from the table.

In the case of /u/, the dependence of both F1 and F2 on acceleration appears to be non-linear: F1 decreases up to 3g then rises again, and F2 increases up to 3g then falls. Given the small number of samples analysed to date, it is not known whether this pattern is significant. Figure 2 shows the positions of the vowels on the F1-F2 chart at 1g and at 5g. Intermediate values are also shown for /u/.

4.3 Variation of F1 and F2 with intra-oral pressure

Table 4 shows the results of measurements of the first two formant frequencies in the lists recorded with positive pressure breathing, at acceleration levels of 4g, 5g and 6g. The nominal pressure of the breathing gas supply is shown in the table. There is evidence that the effect of intra-oral pressure on the cross-sectional area of the pharynx is highly non-linear [6], so no attempt has been made to apply linear regression. Instead, analysis of variance has been used to test for significant differences in formant frequencies between the different conditions, including 5g without pressure breathing. The final column in Table 4 indicates which formants changed significantly (H0: No difference between conditions, $p<5\%$) between the conditions.

The performance of the formant tracker was less reliable on the lists recorded with pressure breathing, because the fundamental frequency F0 was high. Average F0 values increased from 130 Hz at 1g to 175 Hz at 5g without pressure breathing. With pressure breathing, the mean F0 was 280 Hz at 6g, and even exceeded 400 Hz in places. Under these conditions, the formant tracker tended to follow harmonics.

Figure 3 shows the F1-F2 chart of the vowels spoken with pressure breathing, with the 1g safety pressure condition for comparison.

5. DISCUSSION

Although in some cases the number of samples of each vowel is small, the formant frequency measurements show reasonable consistency. At 1g, the values of F1 and F2 are generally similar to the expected values for the general male population. An exception is /u/, which has a much higher F2 than normal; this is similar to previously reported results [1] and may be the result of the pressure of the oxygen mask limiting protrusion of the lips.

The changes in F1 and F2 under acceleration are also generally in agreement with previous work [1], but a much wider range of vowels has been studied here. The high front vowels /i, I, e/ all show F2 reducing as the G level increases, while F1 hardly

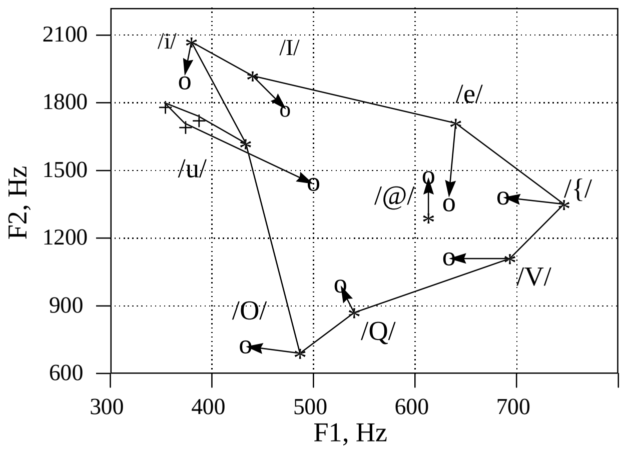

Figure 2 F1-F2 chart showing effects of Gz
* 1g o 5g

Acc.		4g	5g	6g	
P, mmHg		**24**	**34**	**44**	
/@/	**F1**	608	627	588	
	F2	1451	1319	1131	**
/i/	**F1**	415	449	544	**
	F2	1817	1607	1470	**
/I/	**F1**	515	533	548	
	F2	1763	1602	1350	**
/e/	**F1**	582	585	583	
	F2	1550	1515	1436	**
/{/	**F1**	667	604	666	
	F2	1325	1292	1348	**
/V/	**F1**	644	603	597	
	F2	1147	1191	1129	
/Q/	**F1**	534	580	626	**
	F2	1016	938	991	**
/O/	**F1**	471	452	527	**
	F2	910	934	1210	**
/u/	**F1**	411	521	555	
	F2	1686	1540	1381	**

Table 4 Mean values of F1 and F2 with pressure breathing (** significant at p<5%)

changes. The neutral vowel /@/ shows an increase in F2, also with little change in F1. The vowels /O, {/ on the other hand, show reductions in F1 while F2 remains almost constant. /V/ also shows this pattern, although the changes are not statistically significant. The behaviour of F1 and F2 for /u/ is similar to that seen for one speaker in [1], except at 5g. In this case, however, the measured formant frequencies show large variation; F1 ranges from 367 Hz to 711 Hz. While there is a tendency for the variance of all the measurements to increase as the G level rises, this is exceptionally high, so it is not considered a reliable result.
Further work is needed to optimise the formant tracker for the highly stressed conditions, where F0 is high. The effects of positive pressure breathing clearly show the area of the F1-F2 space collapsing in towards a central point. The high vowels are especially strongly affected. Examination of the data also shows that there are trends within each condition. Utterances occurring early in the list are generally further from the central point, while those occurring later are closer to it. For example, F2 for /e/ at 6g/44 mmHg is 1676 Hz in the first utterance, but only 1290 Hz in the tenth. This effect is almost certainly due to fatigue.

6. CONCLUSIONS

Formant frequencies for nine vowels have been measured from recordings made under acceleration of up to 5g, and also under accelerations of 4g, 5g and 6g with positive pressure breathing. To date, the speech analysed has been from one speaker only. The changes in F1 and F2 are in general agreement with previous work in showing a tendency for the vowel space to contract when the speaker is under high acceleration. The present work has studied many more vowels, however, including back vowels for which F1 is reduced, while F2 is unaffected.

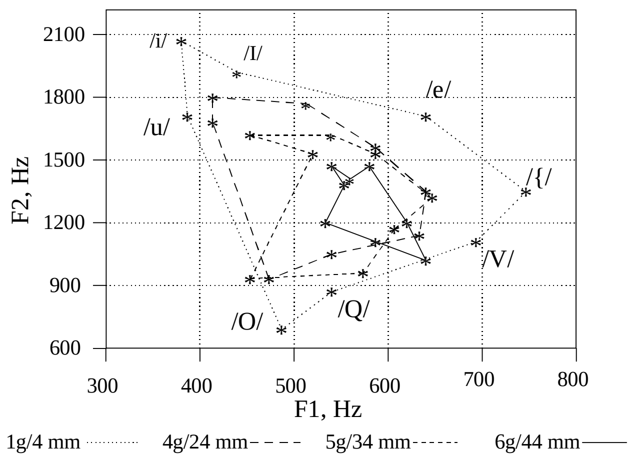

Figure 3 F1-F2 chart showing effect of intra-oral pressure

When positive pressure breathing is introduced, the reduction in the vowel space is drastic: at 6g with a pressure of 44 mmHg, the range of F1 is little more than 100 Hz, and that of F2 only 500 Hz.
Further work will include improving the performance of the formant tracker on the highly stressed lists with a high F0, and measuring formant frequencies in the speech of four other speakers. If the effects are sufficiently consistent, techniques of voice transformation will be investigated, with the aim of normalising the speech before it is applied to a speech recogniser, or coder.

7. REFERENCES

[1] Z S Bond, T J Moore and T R Anderson "The effects of high sustained acceleration on the acoustic-phonetic structure of speech: a preliminary investigation" Armstrong Aerospace Medical Research Laboratory, AAMRL-TR-86-011, 1986

[2] C Gulli, et al, "G-load effects and efficient acoustic parameters for robust speaker recognition" in *Advanced Aircraft Interfaces: the Machine side of the Man-Machine Interface* AGARD CP-521 October 1992

[3] A J South "Voice recognition in adverse aircraft cockpit environments" in *Audio Effectiveness in Aviation* AGARD CP-596, Copenhagen, June 1997.

[4] Unpublished MoD material

[5] Unpublished MoD report

[6] J. Ernsting "Some effects of raised intrapulmonary pressure in man" AGARDOGRAPH 106, Technivision Ltd., Maidenhead, 1966