A Model of Vowel Production under Positive Pressure Breathing

*Allan J South**

Defence Evaluation and Research Agency
Farnborough, Hampshire, GU14 0LX, UK
ajsouth@dera.gov.uk

Abstract

Agile combat aircraft currently in development using speech recognition as part of the cockpit interface will also use positive pressure breathing (PPB) to enable the pilot to remain conscious and function at high G levels. Recogniser performance is seriously affected by PPB even at moderate G levels when the intra-oral pressure is only 10-20 mmHg. This paper describes work which extends the standard n-tube model of vowel production to include the effects of intra-oral pressure. The ultimate aim of this work is to improve speech recogniser performance under these highly stressed conditions.

An 8-tube Dynamic Regions and Modes model is used for this work, with the assumption that all regions of the vocal tract wall have the same compliance. A side branch was added to the model at the junction of the larynx tube and the lower pharynx to simulate the oesophagus. The model shows that as the intra-oral pressure is increased, the vowel space in the F1/F2 plane shrinks towards the region of F1 = 400 Hz, F2 = 1200 Hz. Measurements made on real speech recorded from six subjects at pressures up to 30 mmHg show a similar trend, although the reduction in the range of F2 is less than that indicated by the model. Variation in compliance of different regions of the vocal tract wall seems likely to be the main reason for the differences between the predictions of the model and the measured values.

1. Introduction

The introduction of Automatic Speech Recognition (ASR) into the cockpit of military aircraft has the potential to reduce pilot workload and improve flight safety. However, the severe physical stresses encountered in combat aircraft will affect speech production in parts of the flight envelope, and hence degrade the performance of the ASR equipment. This paper describes results of research into the effects of one such stressor, namely positive pressure breathing.

Positive pressure breathing (PPB) may be used to prevent G-induced loss of consciousness. The pressure of the breathing gas supply to the pilot's oxygen mask is increased in proportion to the G force, above the pressure in the cockpit. This causes the pilot's blood pressure to increase and thus maintains an adequate supply of oxygen to the eyes and brain. The pilot wears a chest counter-pressure jerkin to protect his lungs, but his throat, cheeks, and lower jaw are not covered, so parts of the vocal tract will be distended, resulting in changes in speech production that have a serious effect on speech recogniser performance [1].

The changes induced in the speech spectrum by this mechanism should be reasonably predictable, so there is a possibility that they could be reversed by signal processing techniques to restore the speech to its "normal" spectrum and hence improve the performance of the ASR equipment on the aircraft under high G conditions. Alternatively, normal speech could be modified to provide training material for the recogniser's models, without subjecting the pilot to unnecessary stress.

Section 2 describes a simple model of vowel production, which has been modified to include the effects of intra-oral pressure. The changes in formant frequencies predicted by this model are described. Section 3 describes measurements of formant frequencies made on speech recorded under PPB, and these are compared with the model's predictions in section 4. Conclusions are drawn in section 5.

2. Vocal tract model with intra-oral pressure.

It is well known that a simple model of the vocal tract consisting of a number of uniform tubes of various cross-sectional areas and lengths can reproduce most of the acoustic characteristics of vowels [2, 3]. The effect of an increase in the pressure within the vocal tract can be included in this model by allowing the cross-sectional areas of the tubes to increase as a function of the pressure.

Following initial experiments with a four-tube model, it was decided to use an eight-tube model based on the Dynamic Regions and Modes concept [4]. In this model, the length of each tubelet is a fixed proportion of the overall vocal tract length. The tubelet lengths are L/10, L/15, 2L/15, L/5, L/5, 2L/15, L/15, L/10, where L is the overall vocal tract length. The various tubelets are in approximate correspondence with anatomical regions of the vocal tract, as shown in Figure 1. This feature of the model would be an advantage if different compliance values were to be assigned to regions of the vocal tract, although the results described here assume uniform compliance throughout the vocal tract.

An important feature of the model is the addition of a side branch at the junction of the larynx tube and the lower pharynx, to represent the piriform sinuses and the oesophagus, which also opens as the intra-oral pressure rises (see below).

The model was implemented using the usual electrical transmission line analogy [3], including the effects of wall vibration, and losses due to viscous and thermal effects. The

* Now at: 20/20 Speech Ltd., Malvern Hills Science Park, Malvern, Worcs. WR14 3SZ. A.South@2020speech.com

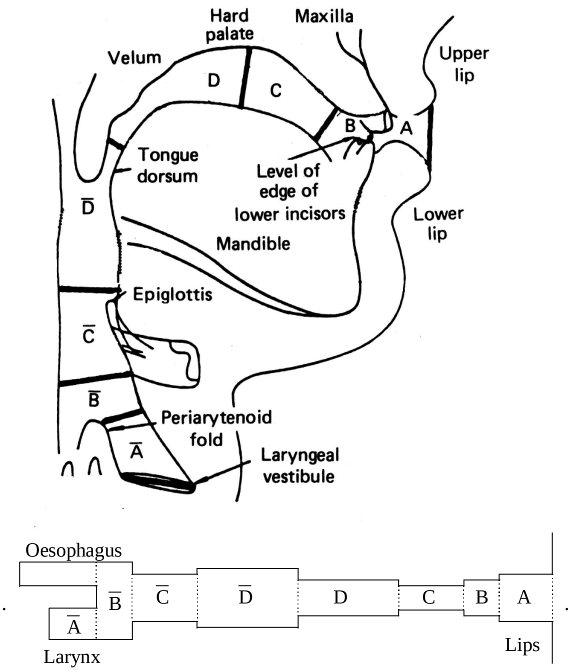

Figure 1 Correspondence of tubes in the DRM model with regions of the vocal tract (adapted from [4])

value of the density of air was also varied in proportion to the pressure.

To account for the effect of intra-oral pressure, it is assumed that the tubelets are circular in cross section and that their radii increase proportionately to the applied pressure. Hence the cross-sectional area of the *n*th tube at pressure *P* will be:

$$A_n(P) = \pi\left(r_{n0} + PC\right)^2$$

where C is the compliance of the vocal tract walls and r_{n0} is the radius of the tube at zero pressure.

The lengths of the tubes are assumed not to change as a result of the increase in intra-oral pressure.

2.1. Compliance of vocal tract walls

For this simplified model, it is assumed that all regions of the vocal tract have the same compliance. It is, of course, recognised that this assumption is likely to be incorrect, but it forms a starting point for the model.

Svirsky et al. [5] reported measurements of the displacement of the tongue surface during bilabial stops, using an electromagnetic articulometer. Compliance values ranging from $0.82.10^{-5}$ to $8.5.10^{-5}$ cm^3/dyne were found, with the lower values for the voiceless stop and higher values for the voiced stop. In part, the range is explained by differences in muscular activity, either actively stiffening the tongue during /p/ or expanding the vocal tract during /b/ in order to maintain a transglottal pressure difference that will allow vocal fold vibration. Svirsky et al. also estimated the average vocal tract compliance during /b/ as $1.94.10^{-5}$ cm^3/dyne.

Some measurements of vocal tract distension were made during investigations of the physiological effects of positive pressure breathing at the Royal Air Force Institute of

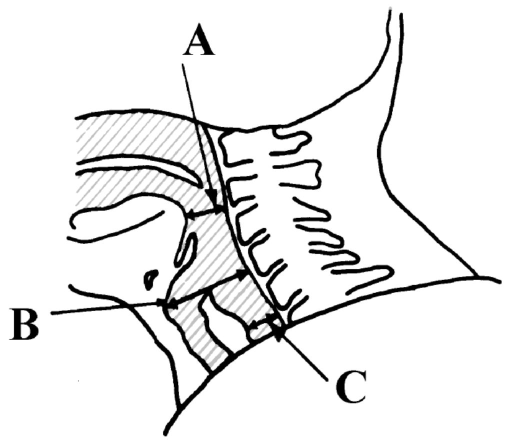

Figure 2 Tracing from radiograph taken during pressure breathing at 80 mmHg (adapted from [6])

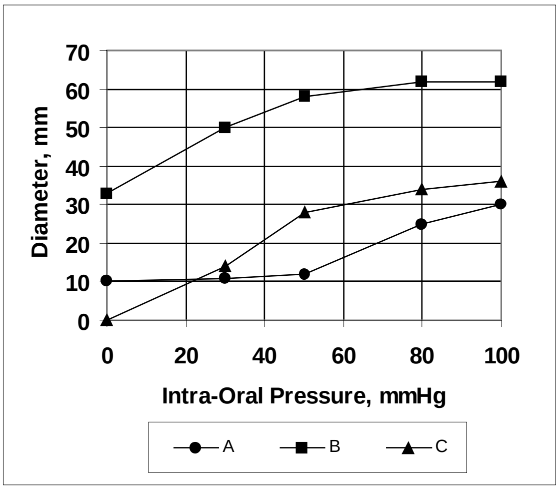

Figure 3 Antero-posterior diameters vs Intra-Oral pressure (adapted from [6])

Aviation Medicine around 1960 [6]. Antero-posterior dimensions of the vocal tract in the upper and lower pharynx for one subject were determined from tracings of radiographs (Figure 2) and are reproduced in Figure 3. No figures were given for lateral distension. The subject was not attempting to speak during this study.

The range of pressure and hence vocal tract distension is very much greater in this work than was considered by Svirsky et al [5]. It is evident from Figure 3 that the distension of the vocal tract is not a linear function of pressure, and that different regions of the tract have different compliance. However, in practice it is not considered likely that speech would be a suitable control modality for functions required by the pilot at G levels at which the breathing pressure would exceed about 30 mmHg. At these lower pressures, the distension of the lower pharynx and oesophagus (curves B and C respectively in Figure 3) is reasonably linear and corresponds to a compliance of about 2.10^{-5} cm^3/dyne. The close correspondence between these two very different studies gives confidence that values in this range can be used

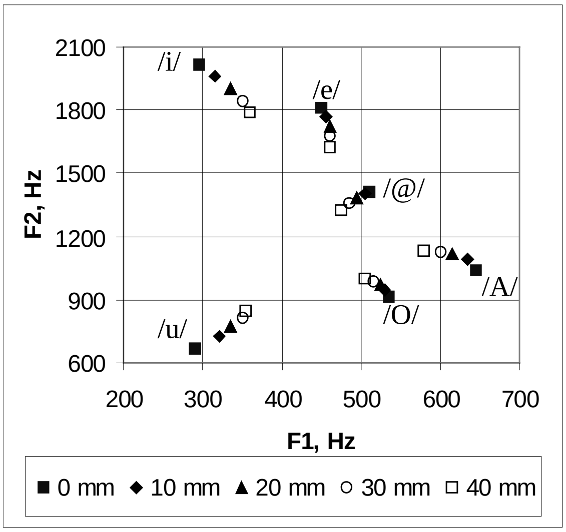

Figure 4 F1/F2 chart predicted by the model for intra-oral pressures up to 40 mmHg

for the model. When a subject is speaking during PPB, it is to be expected that the muscles in the walls of the vocal tract will be tensed, so a value of compliance towards the lower end of the range of measured values should be used.

The upper pharynx shows very little distension at pressures up to 50 mmHg (curve A in Figure 3). No figures are given for lateral distension, although Ernsting comments that "The antero-posterior radiographs show that there was also gross lateral distension of the pharynx."

Ernsting's studies also showed that the upper part of the oesophagus opens as the intra-oral pressure increases (curve C in Figure 3). It is suggested that the opening could be up 8 cm long, but in the lower part of the oesophagus, adjacent to the lungs, the pressure from the chest counter-pressure jerkin will balance the internal pressure, so that the length of the opening is probably limited to about this value. There are no measurements to suggest at what rate the length of the opening increases as the pressure increases.

2.2. Variation of formant frequencies predicted by the model

For each of the vowels /i, e, A, O, u/, the area parameters of the vocal tract model were adjusted to match the vocal tract area functions measured by Fant [2] as closely as possible. The area of each tubelet was adjusted such that its volume was equal to the volume of the corresponding region of the measured area function. The area of the tubelet $\overline{B}$ representing the lower pharynx was then reduced by the volume of the side branch representing the piriform sinuses.

A uniform tube was also included, represented by the schwa /@/ in Figure 4.

From these initial area functions, the model was used to re-calculate the formant frequencies at pressure increments of 10 mmHg up to 40 mmHg, using a value of 10^{-5} cm^3/dyne for the compliance of the vocal tract walls. In addition, the length of the side branch representing the oesophagus was increased at the rate of 0.03 cm/mmHg from an initial value of 1.8 cm. The results for the six vowels are shown in Figure 4, where the outermost points mark the zero pressure results in all cases.

3. Formant frequencies measured under PPB

The database used for this experiment comprised recordings made by six male subjects under conditions of positive pressure breathing alone (i.e. without acceleration) [8]. The original purpose of these recordings was to characterize the effect of PPB on speech recogniser performance. Lists of digit strings and cockpit command phrases were recorded at pressures up to 70 mmHg. During the recordings, the subject wore a flying helmet and oxygen mask, pressure jerkin and anti-G trousers. For the higher pressures, 30 mmHg and above, the jerkin and anti-G trousers were inflated to the same pressure as the breathing gas supply to the mask. At low pressures, up to 30 mmHg, the garments were disconnected. (Hence, 30 mmHg was recorded twice, with and without counter-pressure.)

The subject's speech was recorded through the standard RAF oxygen mask microphone and aircraft intercom system, as the aim of the experiment was to gather data to enable prediction of recognition accuracy in fighter aircraft applications. All speech files were downsampled to 8 kHz, then filtered to compensate for the mask and microphone responses. After pre-emphasis, formant tracks were generated automatically, using proprietary software, from a 14th order LPC analysis in a 20 ms window with 50% overlap. All formant tracks were checked manually at the points of measurement by overlaying them on a spectrogram.

The microphone used has a strong peak in its response at a frequency of about 2700 Hz, and this, combined with resonances of the mask cavity in a similar frequency range, has made reliable measurement of F3 impossible in most cases, in spite of attempts to compensate by filtering. However, F1 and F2 have been measured for the vowels /i, e, {, A, O, u/ recorded at pressures up to 45 mmHg. Between five and ten examples of each vowel were measured for each speaker. Where possible, the samples were taken from identical contexts, but this was not always possible. However, the same selection was used for all speakers.

Preliminary analysis of the results showed that there were significant differences between the formant frequencies of vowels recorded with and without chest counter pressure at 30 mmHg. To maintain comparability, therefore, only results from the conditions without chest counter pressure are described here. The conditions included pressures of 5, 15 and 30 mmHg for all vowels. The list containing the digits in which the vowels /i, O, u/ were measured was also recorded at zero pressure.

These results, averaged across all speakers, are shown in Figure 5 in the form of an F1/F2 chart. Within-class standard deviations were typically 45 Hz for F1 and 100 Hz for F2.

4. Discussion

Comparison of Figure 4 and Figure 5 shows a qualitative agreement between the model and the measured results, in that there is a general tendency for the vowel space in the F1/F2 plane to contract as the intra-oral pressure increases. This may be understood in terms of the sensitivity of the formant frequencies to the dominant constriction in the vocal

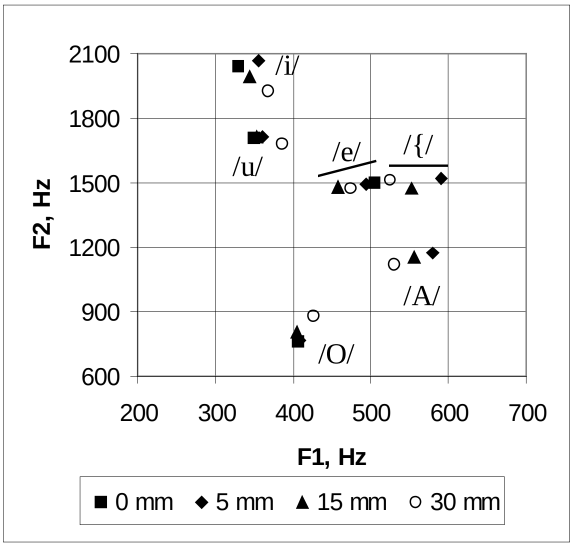

Figure 5 F1/F2 chart of measured values against intra-oral pressure

tract. A fixed expansion of the walls of the vocal tract will change the cross-sectional area in the constriction by a proportionately larger amount than in the wide sections, thus reducing the perturbation of the formant frequencies from the values they would have in a uniform tube.

In the absence of the oesophageal side branch, as the pressure is increased the vocal tract model would tend towards a uniform tube, having formant frequencies at (2n-1)c/4L, i.e. F1 = 515 Hz and F2 = 1545 Hz, if L = 17 cm. The effect of the side branch representing the piriform fossa and the oesophagus is to reduce the frequencies of the lower formants [7]. It can be seen in Figure 4 that the loci of the vowels are tending towards the region around F1 = 400 Hz and F2 = 1200 Hz. Another effect of the side branch is to introduce a zero in the vowel spectrum at a frequency which is dependent on the intra-oral pressure, assuming that the length of the oesophageal opening is affected by the pressure level. This zero is likely to be within the frequency range of interest for speech recognisers. With the model parameters given above, it will be about 2900 Hz when the intra-oral pressure is 40 mmHg.

While the measured results show a similar trend, there are considerable differences in the detail. In the case of the high vowels /i, u/ there is less change in F1 than the model predicts. This also applies to /O/. The low vowel /A/, on the other hand, shows rather more change in F1 than the model predicts.

The change in F2 for the front vowel /i/ is quite close to the model values, but for /O/ the measured change in F2 is very much less than predicted. For /A/, F2 decreases when the model predicts an increase. The articulations used by the subjects in these recordings for the vowels /e, u/ are obviously very different from those of Fant's subject, resulting in quite different positions on the F1/F2 chart. Direct comparisons are therefore impossible, but these two vowels do reinforce the general picture of contraction of the vowel space.

The main reason for the differences between the model and the measured values is likely to be that the assumption of uniform compliance in all regions of the vocal tract is incorrect. This is supported by the observation that the measurements made by Ernsting (1966) actually show very little distension in the upper pharyngeal region for pressures below 45 mmHg (Figure 3). The dominant constriction of the vocal tract for the vowel /O/ is in this region, so the formant frequencies may be expected to change less than predicted by the model, as is actually the case.

5. Conclusions

A simple model has been developed to account for changes in vowel spectra under conditions of positive pressure breathing, using the assumption that all parts of vocal tract expand uniformly as the intra-oral pressure increases. The predictions of this model for a number of vowels have been compared with average values of F1 and F2 for utterances made by six speakers at pressures up to 30 mmHg. The measurements confirm the model's predictions of a reduction in the vowel space in the F1/F2 plane, but the contraction in the F2 dimension seems to be less than predicted. Much of the difference may be due to differences in compliance between various regions of the vocal tract.

6. Acknowledgements

Recording of the speech database used in this experiment was funded by the MoD Applied Research Package 26F, "Air Human Factors". The analysis of the speech was funded in part by DERA Air Systems' Innovation in Research programme.

7. References

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

[2] Fant, G., *The Acoustic Theory of Speech Production* Mouton, The Hague, 1960.

[3] Flanagan, J. L., *Speech Analysis Synthesis and Perception*. Springer-Verlag. New York, 1972

[4] Mrayati, M., Carré, R., Guerin. B., "Distinctive regions and modes: a new theory of speech production" *Speech Communication* 7, October 1988

[5] Svirsky, M. A., et al. "Tongue surface displacement during bilabial stops." *J. Acoust. Soc. Am*er. 102 (1) July 1997

[6] Ernsting, J., *Some effects of raised intra-pulmonary pressure in man.* AGARD Monograph 106, Technivision Ltd., Maidenhead, 1966

[7] Dang, J., Honda, K., "Acoustic characteristics of the piriform fossa in models and humans." *J. Acoust. Soc. Amer.* 101 (1) January 1997

[8] Unpublished MoD Report