

HELIUM SPEECH UNSCRAMBLING: A NEW PERSPECTIVE ON FACTORS AFFECTING INTELLIGIBILITY

G. Duncan[†]

ABSTRACT

The intelligibility of speech uttered in a hyperbaric helium-oxygen (heliox) respiratory mixture is affected in the main by an overall nonlinear frequency translation of the speech spectrum. Spectral resonances (formants) in the short-time spectrum of voiced speech in particular have generally been characterised by a deterministic nonlinear shift curve relating formant centre frequencies in normal air speech to their corresponding frequency locations in helium speech. New results are presented which, whilst supporting the general principle of a nonlinear formant shift in heliox, are the antithesis of the classical theory relating to the formant shift characteristic. It is shown that different speech sounds (phonemes) exhibit characteristic formant shift profiles independent of gas mixture and pressure. These results imply that the helium speech characteristic is affected not only by the properties of the respiratory environment itself, but also by apparently deliberate attempts by the diver to render his own speech intelligible to himself as he perceives it.

INTRODUCTION, SPEECH MATERIAL AND METHOD OF ANALYSIS

Advances in real time signal processing have led to technologically complex electronic systems for unscrambling speech in a helium-oxygen environment at hyperbaric pressures, and yet there has still been cause to vilify the quality of the unscrambled helium speech they produce. There are two corollaries into which the root causes for the poor performance of these systems may be divided: (1) there may be assumptions pertaining to the acoustic events in helium speech which these systems make which are false, or indeed the devices may be deficient in their provisions for processing certain acoustic attributes of the signal which have an important bearing upon intelligibility; and (2) the unscrambling algorithms which are implemented in these devices may affect the resultant unscrambled speech in some manner which is antagonistic to good intelligibility. In this paper, new phenomena are presented from an acoustic analysis of helium speech which support proposition (1) above.

The speech material to be recorded was chosen on the basis that (a) the material should not be linguistically difficult to pronounce and should lend itself to as natural a pronunciation as possible; (b) it should provide a high confidence in terms of constancy of performance of the subject from reading to reading. The material chosen consisted of lists of carrier phrases of the form "I say heed sometimes", with the total list comprising 10 voiced vowel sounds targeted for analysis: /i:/, /o/, /u/, /uh/, /i/, /e/, /a/, /aa/, /uu/ and /@@/ (MRPA). It was possible only to use one volunteer male diver, and recordings were made in a diving simulation hyperbaric chamber, with all lists being spoken once only at, nominally, 100ft depth intervals during the compression phase of the dive, down to a maximum depth of 500ft. The subject had already experienced a hyperbaric heliox atmosphere several times before, but had no phonetic training whatsoever. All readings, including the reading in air at surface pressures, were made within the chamber, with the subject standing in a free atmosphere, unencumbered by personal breathing apparatus. The effective recording bandwidth both for air and heliox was from 20Hz to 10kHz. The first 3 formants of speech

[†]Centre for Speech Technology Research, University of Edinburgh, 80 South Bridge, Edinburgh EH1 1HN, Scotland.

in air occur below 3.5kHz and for a worst-case 100% helium environment the expected linear frequency shift is simply the ratio of the velocity of sound in helium, c_h with respect to that in air, c_a , i.e. $c_h/c_a \approx 3$. Thus, the first 3 formants of speech recorded in the heliox environments of this experiment should be located below 10kHz, corresponding to the bandwidth of the recording system. Speech data was digitised, inclusive of necessary anti-aliasing requirements, to 12-bit resolution. The provision of formant centre frequency data is based here on the parametric spectral estimation technique of autocorrelation-based linear predictive coding (LPC) analysis. The speech data for analysis here is contaminated to some extent by a noise source, located within the diving chamber, due to equipment for cleansing excess carbon dioxide from the respiratory mixture. Since the noise characteristic may affect the spectral estimate by introducing extraneous spectral peaks, then a moderate LPC model order, $p=14$, was chosen for the analysis. This represents a reasonable compromise since too large a model order would produce spectral estimates cluttered by extraneous peaks relating to the noise. However, given a signal-to-noise power ratio of 15dB or better in most recordings, then this model order will model the most intense spectral resonances first, i.e. those relating to the vocal tract formants. A preemphasis factor of $\mu=0.98$ was applied to all segments prior to analysis, and an autocorrelation (analysis window) length of 25.6ms was chosen. Tracking of the individual formants of each vowel from depth to depth was performed manually by visual identification of formant peaks, but evaluation of formant centre frequency was determined algorithmically by a peak-searching routine which approximates resonance features by a parabolic curve. Lastly, in an effort to mitigate the fluctuations in formant frequency which inevitably occur over the duration of the voiced speech segment, several contiguous LPC spectra were added together for each vowel sound to produce one representative average spectrum to which the peak-picking algorithm described above was applied. The spectra from 9 contiguous frames in all were summed and averaged in the analysis of each vowel at each depth, with a spacing between successive spectral estimates of 3.2ms. Thus, the overall time of analysis for each vowel segment is $(25.6 + 8 \times 3.2) = 51.2$ ms. This is admissible in the vowel sections analysed since they relate to the group of voiced monophthongs, which are considered to be among the more steady-state speech sounds, demonstrating long-term stability of formant frequencies and being of long duration (> 70 ms). The averaging of spectral frames in this way has the additional advantage of attenuating the effects of additive Gaussian noise (to which the major additive noise source can be approximated) on the spectral estimate.

FORMANT SHIFT IN HYPERBARIC HELIOX: A NEW PERSPECTIVE

Early research (ref.1) produced a deterministic equation relating formant frequency in heliox, f_h to corresponding formant frequency in air, f_a , which grossly depended on physical properties of the gas such as ratio of specific heats, γ , pressure, P and velocity of sound, c . That is:

$$f_h = \frac{c_h}{c_a} \sqrt{f_a^2 + \left(\frac{c_a^2 \gamma_h P_h}{c_h^2 \gamma_a P_a} - 1 \right) f_{wa}^2} \quad (1)$$

where f_{wa} is the closed vocal tract resonance frequency in air, and is ≈ 180 Hz. Basically, as the density of the gas increases with ambient pressure, then the tissue of the vocal tract becomes less of a perfect acoustic reflector and begins to absorb energy, causing the tissue itself to resonate. Helium formant frequencies for F_1 to F_3 for all 10 vowels analysed at a depth of 400ft ($P=12.96$ bar, 97.5%He, 2.5% O_2 , $c_h/c_a=2.7$) are shown plotted against their formant frequency in air in fig.1, with the theoretical characteristic of equ.(1) demonstrated by the solid line. Note the nonlinear nature of the curve at low frequencies. It can be seen that, although the formant frequencies are roughly distributed as determined by equ.(1),

there are nonetheless large deviations from the characteristic.

Fig. 2 demonstrates a novel reinterpretation of the relationship between formant frequencies in heliox and in air. The heliox formants at each depth for a single vowel phoneme (/ii/) have here been collated and plotted against formant frequency in air to form one new graph. The data for F_1 , F_2 and F_3 corresponding to any given depth have been coded by the same letter (e.g. 'B' identifies values for F_{1-3} at 200ft depth); tracing of the letter code from F_1 to F_2 to F_3 provides a characteristic formant shift profile for each depth, and reveals a significant trend: each formant profile demonstrates a similar characteristic shape irrespective of depth, that is, composition of the heliox respiratory mixture. Moreover, the profile is of a different shape to the predicted quadratic relationship of equ.(1). A similar phenomenon has been observed in the case of each vowel analysed, with the formant data at each depth for any one vowel producing a characteristic profile. Thus in fig.(2) and figs.3(a-c), each solid line joins the mean values of F_1 , F_2 and F_3 averaged over all depths, and represents the characteristic formant shift profile for each vowel *independent of depth*. It can be seen that there is an important disparity in profile shape from one vowel phoneme to another. Indeed, the results of these graphs imply that each vowel has a tendency to produce an individual formant frequency shift signature in a hyperbaric heliox respiratory mixture.

DISCUSSION

A possible cause for this effect is that (a) the subject may be trying to consciously alter his speech in a consistent manner for each phoneme so as to palliate the effect of formant frequency translation due to the physical properties of the heliox gas mixture. Since each speech sound involves different articulatory postures in its production, then this may explain the different profiles for each phoneme and the consistency of profile shape with depth. It is also possible (b) that the helium speech waveform may in itself be acceptable to the talker, but that the feedback of his own voice through the heliox environment may induce him to make changes in his speech output. The reasoning offered here is that that the arrival time at the ear of the speech waveform transmitted through the heliox atmosphere will be different compared to that of air, whereas the arrival time of the signal transmitted by, for example, bone conduction will remain unchanged. The talker may therefore be adapting his speech to minimise the effect of the unusual combination of speech waveforms presented to his own perceptual system. Which, if any, of the above causes is responsible for producing the formant profiles has important implications in terms of helium speech unscrambler architecture. In particular, if the diver is indeed adapting his own speech to attempt to restore formant frequency ratios to their values in air, then the inclusion of speech recognition technology becomes an important factor in successful unscrambling of helium speech, since the perceptual mechanism of each diver is likely to adapt differently, hence denying the application of a single overall formant shift criterion such as defined in equ.(1).

ACKNOWLEDGEMENT

The author wishes to express his kind thanks to Mr. B. Thomson, diver, Subsea Offshore Ltd., and Dr. L. Virr, AMTE/EDU Portsmouth. This research has been supported by the Procurement Executive, Ministry of Defence.

REFERENCES

1. G. Fant and J. Lindqvist, "Pressure and gas mixture effects on divers' speech," *Quarterly Progress and Status Report, Speech Transmission Laboratory STL-QPSR-1* pp. 1 - 17 Royal Institute of Technology, (1968).

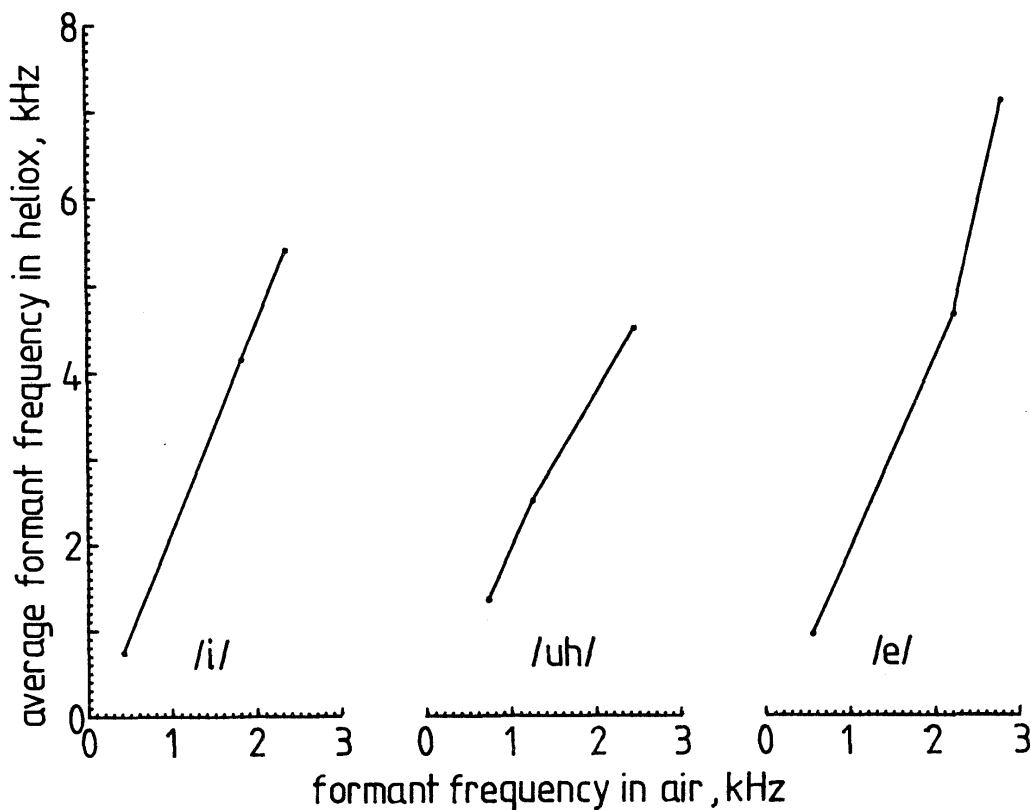
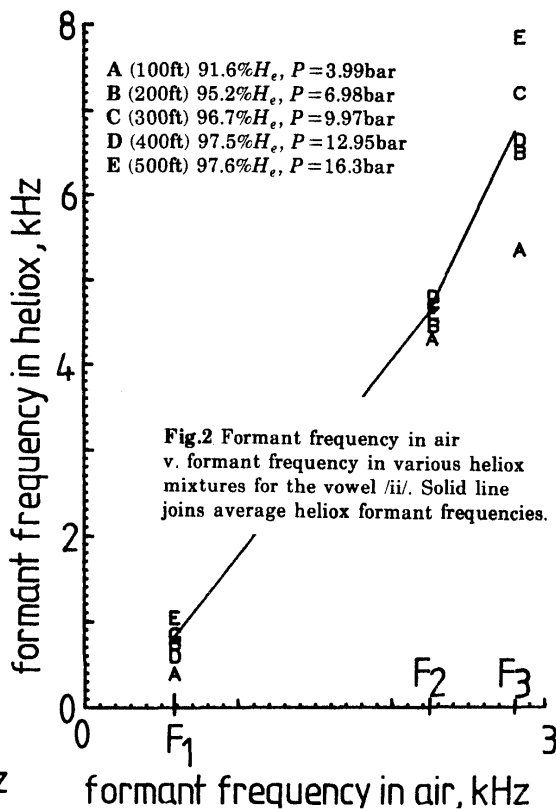
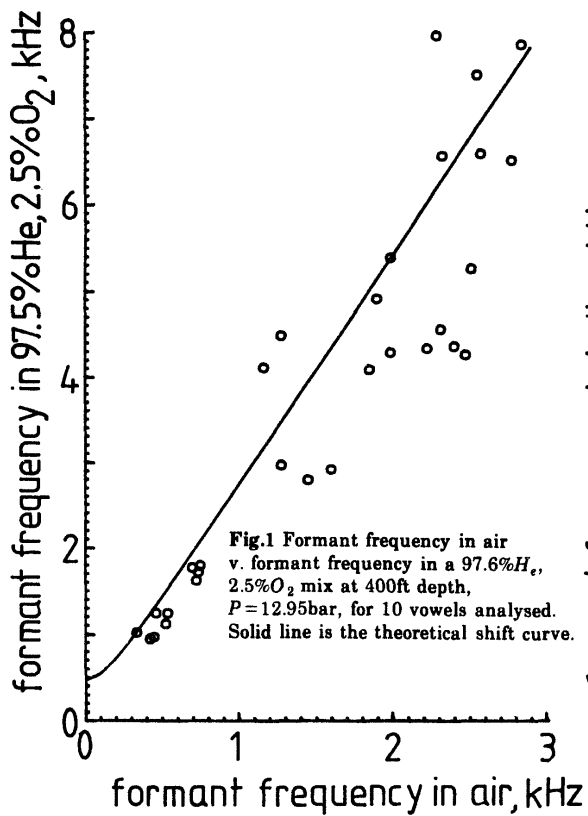


Fig.3 Formant shift profiles (formant frequency in air v. average formant frequency in heliox) for the vowels /i/, /uh/ and /e/.