A SPECTRUM-BASED LPC PROCESSOR FOR ENHANCING THE INTELLIGIBILITY OF HELIUM SPEECH

G. Duncan and M.A. Jack

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

RESUME

Le soulagement de gênes physiologiques dans la plongée profonde exige le remplacement des mélanges oxygène-azote de l'air normal par des melanges binaires comportant un gaz plus léger (d'habitude l'hélium) mais qui entraîne de nouveaux problèmes au niveau de la déformation de la voix, dont l'intélligibilité est beaucoup degradée. Les résonances spectrales de la voix (formants) sont étalées non linéairement par les propriétés physiques du gaz sous pression, mais la fréquence fondamentale des cordes vocales n'est pas affectée par l'atmosphère.

Le systême detaillé dans cet article permet la manipulation de la parole en hélium afin d'améliorer l'intélligibilité en employant la modélisation autorégressive du signal. La structure temporelle du signal est totalement conservée, mais les formants peuvent être comprimés non linéairement afin d'offrir une bonne correction pour la déformation vocale.

SUMMARY

The alleviation of physiological discomfort in deep-ocean saturated diving operations demands the use of helium-oxygen (heliox) breathing mixtures containing large amounts of helium gas, whose effect on the speech waveform is to shift spectral resonances (formants) by a factor nominally equal to the ratio of the velocity of sound in heliox to that in air, although the actual spectral shift is nonlinear, resulting in the degraded intelligibility of helium speech. However, fundamental frequency of vocal tract excitation by the vocal cords remains unaffected.

The residually excited linear predictive coding unscrambler system detailed here permits manipulation of the helium speech waveform to restore intelligibility by use of autoregressive signal modelling, in which the temporal features of the speech signal are totally conserved, but formant data can be corrected nonlinearly.

1 - INTRODUCTION

In hyperbaric deep-ocean diving environments, the use of lighter-than-air respiratory mixtures is of great physiological benefit to the diver. Helium-oxygen (heliox) mixtures greatly alleviate the effects of nitrogen narcosis and dyspnea (respiratory difficulty) which otherwise occur in a high-pressure air environment. However, the principal disadvantage of a heliox mixture relates to its disastrous effect upon voice communications. The heliox mixture, with its increased velocity of sound and different acoustic impedance with respect to air, alters the speech uttered by the diver to such an extent that its intelligibility is heavily impaired. Consequently, serious and occasionally fatal mistakes in comprehension can occur if communication depends on the perception of raw helium speech.

The physical properties of heliox change the centre frequencies and bandwidths of any resonating cavities filled with the gas compared to their values in air, scaling such values by a factor $1<K\leqslant 2.931$, where $K = c_h/c_a$, c_h = velocity of sound in heliox, and c_a = velocity of sound in air, and K = 2.931 for 100% helium. The compounding effects of a high-pressure atmosphere produce a nonlinear upward shift of the frequency response of the human vocal tract[1], although the exact nature of the nonlinearity is not well understood. It is however generally accepted that vocal tract resonant (formant) energy in the helium speech spectrum is attenuated by -6dB for every octave shift upwards in frequency. Additionally, temporal and spectral characteristics of the glottal excitation waveform which enters the vocal tract, including any periodicity of the waveform, are totally conserved from air to helium speech. Thus, viewing the speech mechanism as an impulse or white-noise excited filter, helium speech unscrambler systems require to apply nonlinear correction uniquely to the vocal tract filter frequency response.

Although many real-time unscrambler systems have been designed based on the simple strategy of time-domain pitch-synchronous waveform expansion, thereby acheiving a gross linear compression of the vocal tract frequency response, these techniques offer no possibility of nonlinear formant correction. They also exhibit discontinuity of speech output when no definite pitch period is present, as in unvoiced speech (e.g. "s" or "f" sounds). Continuity of speech is achieved by a frequency-domain approach to unscrambling. In using the short-time Fourier transform (STFT)[2], techniques are available for conservation of the periodicity information in the composite spectrum of a short-time segment of speech. The problem is then one of separating the glottal excitation spectrum from the vocal tract frequency response so that correction for the helium speech effect can uniquely be applied to the latter. Corrected vocal tract and conserved glottal excitation estimates are then recombined and inverse Fourier-transformed to form a new corrected short-time air speech segment. However, in addition to the complexity of requiring an overlap-and-add (OLA) approach[3], due to requirements of windowing to reduce spectral leakage, the principle enigma of the STFT approach to unscrambling relates to spectral phase, which is left intact in all correction operations under the pretext of the phase insensitivity of human hearing. Nonlinear correction is therefore confined to magnitude spectra only, but subsequent recombination with an inappropriate phase spectrum can be shown to cause distortions which affect the intelligibility of the resultant speech[4].

Cepstrum-based unscrambling[5] avoids the complexity of an OLA approach since, although time-domain windowing is still necessary, the cepstrum method seeks to construct a finite impulse response (FIR) filter representing the corrected vocal tract which is then reconvolved in the time-domain with an estimate of the glottal excitation waveform. Windowing therefore has the desirable property of improving the estimate of the FIR filter by reduction of spectral leakage, rather than affecting any composite speech spectrum. However, the technique still suffers from essentially unknown phase characteristics both for the estimated helium vocal tract cepstrum and in the corrected spectrum.

2 - A SPECTRUM-BASED LPC PROCESSOR FOR RESTORING INTELLIGIBILITY TO HELIUM SPEECH

The novel method presented in this paper employs LPC-based vocal tract estimation and glottal excitation deconvolution by inverse filtering, but applies explicit spectral-domain helium speech unscrambling permitting nonlinear correction to short-time autoregressive portions of the speech signal [6] . The glottal waveform, which must ultimately be totally conserved, is estimated by calculating inverse filter - or prediction error filter (p.e.f) - parameters and applying each p.e.f to its corresponding short-time segment of helium speech. However, this system is unique in its approach to helium speech unscrambling by its use of the Weiner-Xinchine theorem in estimating both heliox and air-equivalent filter structures.

The solution for the coefficient set a_k of the p.e.f using least mean squares techniques necessitates the calculation of the autocorrelation sequence of the signal. From the Weiner-Kinchine theorem:

$$R(\tau) = \int_{-\infty}^{\infty} P(f) e^{j2\pi\tau} df \qquad (1)$$

where $P(f)$ is the power spectrum of the signal and $R(\tau)$ is the autocorrelation sequence. For a real-valued signal such as speech, then the power spectrum can simply be calculated by applying a short-time Fourier transform to the signal and squaring and adding real and imaginary frequency components at each discrete frequency. Calculation of the helium speech autocorrelation function $R_h(\tau)$ is then possible by applying the transform of equ.(1) above. Note however that special precautions are necessary in digital signal processing to force a linear autocorrelation from the cyclic autocorrelation afforded by the discrete Fourier transform [7] .

The strategy of using the power spectrum as a route to the autocorrelation function is particularly convenient since the power spectrum implicitly contains spectral information relating to vocal tract formants and therefore can itself be directly corrected for the helium speech effect. The resulting power spectrum $P_a(f)$ corresponding to normal air speech can be used to form the autocorrelation function $R_a(\tau)$ of the spectrally corrected signal, permitting the calculation of an air-equivalent inverse vocal tract p.e.f sequence, b_k , say, from which a resynthesis filter is easily constructed and excited by the conserved residual excitation. Note that in this method, the power spectrum, prior to spectral correction, still contains information regarding the glottal excitation, since no attempt is made to deconvolve this when applying the Fourier transform to the short-time segment. Therefore applying spectral compression to the power spectrum of voiced speech necessarily entails an effective increase in the fundamental frequency of the resulting speech. However, this apparent imperfection is redressed here by a consideration of an implicit but rarely explicitly-expressed property of autoregressive signal processing. Namely, although any given p.e.f is constructed from a signal having, say, a well-defined fundamental frequency, the resulting inverse filter will produce a spectrally-white residual when applied to any similar signal originating from the same linear time-invariant (LTI) system, but which possesses a different excitation periodicity. A general overview of the spectrum-based helium speech unscrambler is shown in figure 1.

3 - FACTORS AFFECTING SPECTRUM-BASED LPC HELIUM SPEECH UNSCRAMBLER PERFORMANCE

This category of unscrambler implicitly offers continuity of the unscrambled speech waveform. In autoregressive modelling in general, the optimum fit to the vocal tract transfer function assumes an all-pole model. Therefore, those features of the signal spectrum which are directly due to the influence of transfer function zeros, i.e. spectral antiresonances, will not be faithfully represented. This is in contrast to STFT and cepstrum techniques, which make no assumptions other than that the system under analysis is LTI in the short term, and which therefore treat all components (poles and zeros) of the vocal tract transfer function. However, although the characteristic transfer function of certain sounds, such as nasals, does indeed contain zeros of transmission, the important aspects of speech perception still relate to formant information, that is, the relative locations, amplitudes and bandwidths of spectral peaks, so vindicating the use of AR modelling techniques. However, whereas this is acceptable in an analysis-only scenario, the explicit application of the p.e.f to the helium speech is unlikely to result in a residual which is spectrally white. The estimate of the glottal waveform will thus contain elements of the short-time spectrum which were not well-matched by the all-pole model, specifically, areas of antiresonance.

However, the spectrum-based LPC unscrambler presented here offers ease of nonlinear correction of all aspects of the short-time vocal tract frequency response. Although the time waveform requires windowing to avoid spectral leakage, no overlap-and-add approach need be applied to the composite output waveform since the spectrum is used to estimate a filter impulse response which then operates on *unwindowed* versions of both the composite input helium speech waveform and the glottal excitation estimate. Preemphasis, which imparts a +6dB/octave emphasis to the input speech spectrum, is a well-established requirement to increase the accuracy of the AR modelling process, and here offers implicit correction for the high-frequency attenuation of helium speech, so obviating any explicit provision for frequency-dependent formant amplitude correction. A matching deemphasis filter cascaded to the synthesis filter output restores the usual air speech formant charcteristics.

Precautions necessary to ensure linear autocorrelation from the inverse power spectrum [7] require that the input time sequence be doubled in length by appending the same number of zeros to the end of the original N-length data sequence. This effectively doubles the power spectrum resolution and hence improves the accuracy of spectral correction, which depends on an index-remapping and interpolation procedure. Note that areas of the power spectrum are undefined after corrective spectral compression, between the frequencies $f_s/2.K \rightarrow f_s/2$ and its image area $f_s/2 \rightarrow (1-1/2.K)f_s$, where f_s is the sample frequency. From a consideration of the maximum entropy properties of AR signal processing, the solution as to the power values to assign to these frequencies consists of simply repeating the last known frequency value (which must be nonzero) in the power spectrum over the undefined region. This is in contrast to the STFT and cepstrum methods, which require spectral tapering linearly to zero across such undefined spectral areas.

Most importantly, the use of the co-phase power spectrum and the Weiner-Kinchine relationship avoids any explicit consideration of spectral phase whatsoever.

4 - EXPERIMENTAL RESULTS

An example of the application of the spectrum-based LPC helium speech unscrambler is demonstrated in fig.2. Figure 2(a) is the power spectrum of a 51.2ms segment of the vowel "ee" extracted from continuous helium speech, sampled at f_s = 20kHz, spoken by a male diver at a depth of 250ft (pressure = 8.5Bar) in an atmosphere consisting of 96% He and 4% O_2. Shown inset is the Hamming-windowed waveform segment for analysis. Note the fundamental excitation periodicity of approx. 6ms, which is also apparent in the power spectrum due to the approx. 167Hz spacing of spectral lines. The vocal tract frequency response is estimated by the dotted-line envelope which has been overlaid on the spectrum, and formants F1, F2 and F3 have been identified.

The power spectrum of the same speech segment corrected for the helium speech effect as output from the unscrambler system is shown in fig.2(b). Note that fundamental frequency has been absolutely conserved. Close examination of the spectra also demonstrates that much of the fine spectral detail relating to the glottal excitation waveform has also been conserved. Each helium formant has been successfully relocated in frequency by a factor of K = 2.4, with an attendant decrease in bandwidth. Note that the output signal has been downsampled by a factor of x2 so that the output sample frequency, f_{so} = 10kHz.

ACKNOWLEDGEMENT

The authors wish to express their kind thanks to Dr. L.E. Virr, Experimental Diving Unit, H.M.S. Vernon site, Portsmouth, for his support in this work. This research has been supported by the Procurement Executive, Ministry of Defence (U.K.).

REFERENCES

1 Jack, M.A., and Duncan, G., "The helium speech effect and electronic techniques for enhancing intelligibility in a helium-oxygen environment", *The Radio and Electronic Engineer*, v.52, n.5, pp.211 - 223 (May 1982)

2 Richards, M.A., "Helium speech enhancement using the short-time Fourier transform", *I.E.E.E. Transactions*, v.ASSP-30, n.6, pp.841 - 853 (Dec. 1982)

3 Crochiere, R.E., "A weighted overlap-and-add method of short-time Fourier analysis/synthesis", *I.E.E.E. Transactions*, v.ASSP-28, pp.99 - 102 (Feb. 1980)

4 de Boer, E., "A note on phase distortion in hearing", *Acustica*, v.11, pp.182 - 184 (1961)

5 Quick, R.F., "Helium speech translation using homomorphic techniques", *Journal of the Acoustical Society of America*, v.48, p.130(A) (1970)

6 Duncan, G., and Jack, M.A., "Residually excited LPC processor for enhancing helium speech intelligibility", *Electronics Letters*, v.19, n.18, pp.710 - 711 (Sept. 1983)

7 Cooley, J.W., Lewis, P.A., and Welch, P.D., "Applications of the fast Fourier transform to computation of Fourier integrals, Fourier series, and convolution integrals", *I.E.E.E. Transactions*, v.AU-15 , n.2, pp.79 - 84 (June 1967)

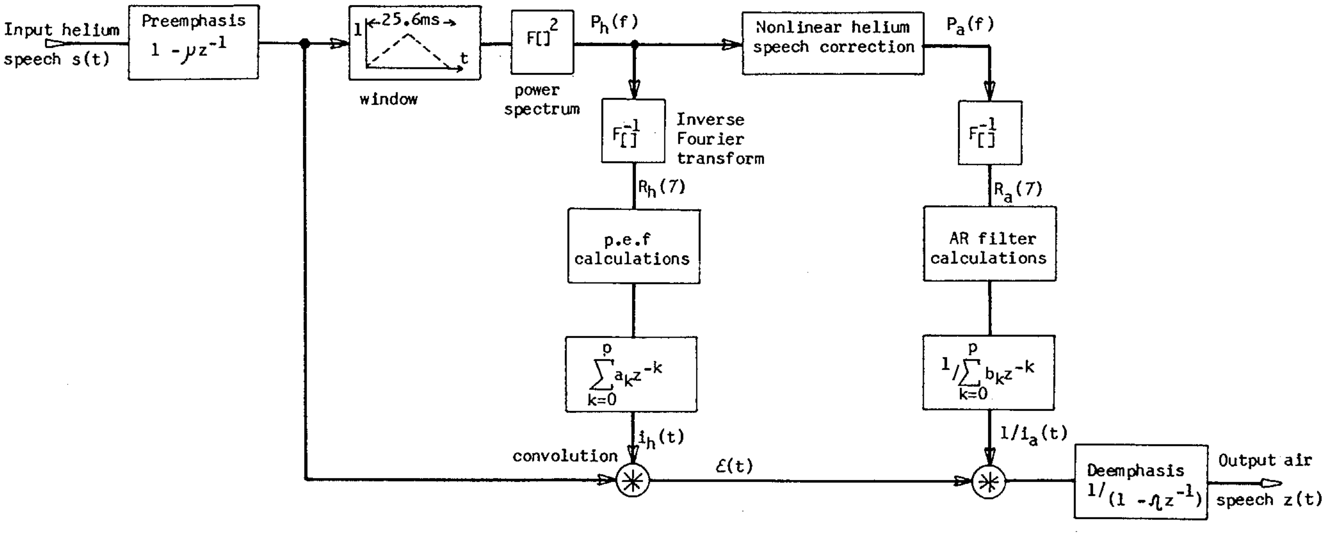
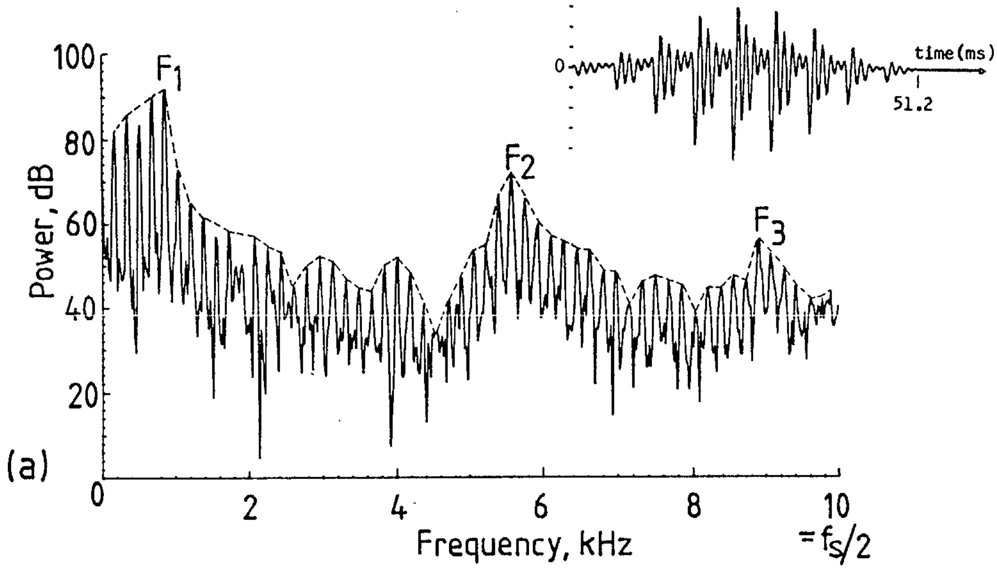


Fig.1 Block diagram of the spectrum-based LPC helium speech unscrambler system. F[] is the Fourier transform operator.



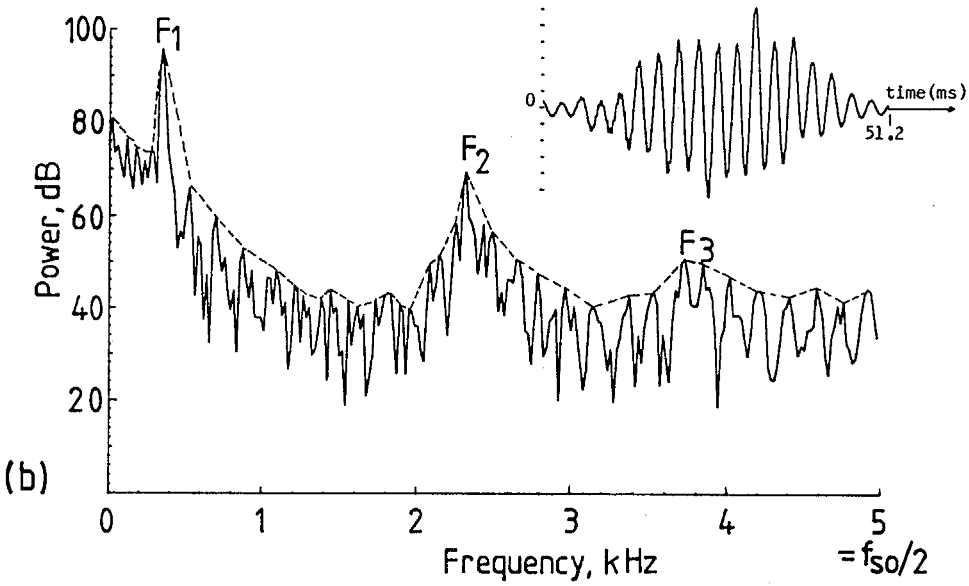


Fig.2 (a) Helium speech power spectral density for the vowel /ee/ (51.2ms) before processing by the spectrum-based LPC unscrambler system.
(b) Power spectral density of corrected speech segment using a linear compression ratio K=2.4.
(Note. The power spectrum of 2(b) is obtained from a post-processing analysis of the output time waveform, z(t)).