Industry
Canada
CRC

HELIUM SPEECH DESCRAMBLER FINAL REPORT (U)

by

Karen Bryden, Hisham Hassanein
and François Marceau

CRC REPORT NO. 95-004

June 1995
Ottawa

The work described in this document was sponsored by the Department of National Defence under Task 041LA.

Canada

Helium Speech Descrambler Final Report (U)

by

Karen Bryden, H. Hassanein
and François Marceau
Advanced Radio Systems Group
Radiocommunications Technologies Research

The work described in this document was sponsored by the Department of National Defence under Task 041LA.

COMMUNICATION RESEARCH CENTRE, INDUSTRY CANADA
CRC REPORT NO. 95-004

Canada

June 1995
Ottawa

ABSTRACT

The presence of helium in the breathing gas used in certain military divingoperations renders the divers' speech unintelligible. This distortionrepresents a danger to the divers. A new, lower cost device was developedto "descramble" the divers speech. The device is based on linearprediction and uses interpolation of the autocorrelation function to removethe nonlinear frequency distortion. It is implemented on a digital signal processing chip in both fixed and floating point arithmetic. Another algorithm based on frequency domain manipulation is described. During tests, it became apparent that distortion caused by the diver's face mask and limitations of the microphone in the diver's helmet were major contributors to communication problems.

RESUMÉ

La présence d'hélium dans le mélange de gaz en plongée sous-marine dans certaines applications militaires rend la parole des plongeurs incompréhensible. Cette distorsion de la parole représente un danger aux plongeurs. Un nouvel appareil à prix modique a été développé pour restaurer la parole des plongeurs. Cet appareil est basé sur la prédiction linéaire et utilise l'interpolation de la fonction d'autocorrélation pour éliminer la distorsion spectrale non-linéaire. L'algorithme a été exécuté en point flottant et point fixe sur un microprocesseur de traitement numérique. Un autre algorithme basé sur la manipulation du domaine spectral est décrit. Durant des tests subjectifs, il est devenu évident que la distorsion causée par le casque de plongée et les limitations du microphone à l'intérieur du casque contribuent d'une façon majeure aux problèmes de communications.

EXECUTIVE SUMMARY

During diving operations, Canadian Navy divers are kept in constant contact with the surface tender via communication lines bundled with the breathing air hoses. Under certain circumstances where depth prevents the use of compressed air as a breathing gas, a mixture of helium and oxygen is used, consequently rendering the divers' speech unintelligible. A helium speech "descrambler" is used to remove as much of the distortion as possible.

The Defence and Civil Institute of Environmental Medicine (DCIEM) tasked the Communications Research Centre (CRC) to investigate the problem and design a new descrambler to replace the aging and ineffective units currently in use. Two new algorithms were designed, built and tested, both in the hyperbaric chamber at DCIEM and during helium test dives held at the Fleet Diving Unit (Atlantic) in Shearwater NS.

Although the new descrambler performed exceptionally well on helium speech recorded without a face mask, neither the new unit or the unit it is intended to replace worked well under realistic conditions, in water and with face masks. This led to a recommendation to DCIEM to investigate distortion caused by the face mask and by the diver microphone, both of which are aggravated by the frequency shifts in helium speech.

The report develops in detail the mathematical foundation for the new algorithm and the design of the new descrambler. An alternative frequency domain approach to helium speech descrambling is also discussed.

INDEX

	Page
Abstract/Resumé	iii
Executive Summary	v
Index	vii
1.0 Introduction and Background	1-1
2.0 Problem Statement	2-1
3.0 Solution Implemented by CRC	3-1
4.0 Descrambler Description	4-1
5.0 Analysis and Test Results	5-1
6.0 Conclusions and Recommendations	6-1
7.0 References	7-1
Appendix 1, Non-integer frequency shifts obtained by linear interpolation of the autocorrelation function	A-1-1
Appendix 2, Discussion of M-101/AIC Microphone and audio system frequency response for diver communications	A-2-1
Appendix 3, Cubic Interpolation of the autocorrelation function	A-3-1
Appendix 4, Frequency domain approach to frequency scaling	A-4-1

1.0 INTRODUCTION AND BACKGROUND

During diving operations, Canadian Navy divers are kept in constant contact with the surface tender via communication lines bundled with the breathing air hoses. Under certain circumstances where depth prevents the use of compressed air as a breathing gas, a mixture of helium and oxygen is used, consequently rendering the divers' speech unintelligible. A helium speech "descrambler" is used to remove as much of the distortion as possible.

A helium speech descrambler made by Hele Co. is used by the Canadian Navy. This unit is expensive, and is often found to operate ineffectively. Other available descramblers such as those made by Stocktronics are extremely expensive. In order to provide a low cost alternative to these descramblers, the Defence and Civil Institute of Environmental Medicine (DCIEM) tasked the Communications Research Centre (CRC) to investigate the problem and design a new descrambler.

CRC had previously simulated a new descrambler and used it successfully to descramble a sample of helium speech recorded at the Naval Ocean System Centre (NOSC) in San Diego Ca. This algorithm was transferred to a digital signal processing (DSP) chip and tested during a set of helium dives performed at the Fleet Diving Unit (Atlantic) in Shearwater NS in Sept. 1993. The Model 1 unit performed as well as, and in some cases better than the Hele unit, in spite of its simplicity.

Recordings made during the trials were brought back to Ottawa and used during the development of a Model 2 descrambler. In Jan. 1994, during subsequent helium chamber dives at DCIEM, divers were asked to speak a set of isolated words from a subjective listening test at representative depths. The divers speech was recorded and later passed through both the Hele descrambler and the Model 2. A comparison of the two descramblers was obtained by playing back the descrambled speech to test subjects who tried to identify each word among a set of similar sounding words. This report details the results of the subjective tests and the algorithms, hardware and software developed by CRC.

2.0 **PROBLEM STATEMENT**

Distortion of Divers' Speech by Helium in Breathing Gas

In modern diving practice, military, commercial and some sport divers [1] routinely dive to depths exceeding 200 feet of sea water (fsw). However the air we breathe at the surface, consisting of approximately 21% oxygen and 79% nitrogen, is not suitable for deep diving. Beyond roughly 130 fsw, the nitrogen in air becomes narcotic and represents a danger to the diver. In addition, air is absorbed by the diver's tissues in proportion to the duration and depth of the dive. Excess oxygen is consumed by the divers metabolism, however the excess nitrogen must be exhaled by the diver before he/she returns to the surface. Failure to expel enough nitrogen during the ascent can result in the formation of bubbles in the blood and tissues leading to decompression sickness, a condition which can be extremely painful and potentially fatal.

In order to increase the safety of the diver, helium is substituted for some or all of the nitrogen. Helium is non-toxic, and is less soluble in the bloodstream. However heliox, a breathing gas composed of helium and oxygen, presents other dangers to the diver. Heliox conducts heat from the body roughly four times faster than air. In addition, the speech of the diver is rendered virtually unintelligible, creating psychological stress and communications problems. The inability of a diver to communicate in an unexpected or hazardous situation can place the diver in extreme danger.

In closed circuit breathing systems, oxygen is metered into the breathing gas at a partial pressure of roughly 1.4 to 1.6 atmospheres (atm). The remainder between this and the ambient pressure (7 atm at 200 fsw) is made up by the helium diluent. As the diver descends, the proportion of helium in the breathing gas increases, and his/her speech becomes more distorted. Part of the reason for this is that sound travels faster in helium. The resonances in the vocal tract (formants) which give rise to the different speech sounds are shifted up in frequency in proportion to the relative amount of helium, resulting in a "Donald Duck" effect which increases with depth.

The velocity of sound in a gas is given by the equation

$$\upsilon = \mathrm{sqrt}\ (\gamma * P / \rho) \qquad ...1$$

where P is gas pressure
ρ is gas density (atomic weight of helium 2, air about 29)
γ is the gas pressure/volume gradient (5/3 for helium, 7/5 for air)

From this equation, we find that sound velocity in pure helium is roughly 4 times faster than in air. In surface supplied diving operations, the ratio of oxygen to helium is 16% oxygen/84% helium. For this mixture, the sound velocity should be 2.3 times faster than in air.

Since the ratio of helium to oxygen is constant for surface supplied dives, one would not expect the frequency shift to vary since from Boyle's law, P / ρ is constant. However lower frequency shifts have been observed at shallower depths, which can perhaps be attributed to the presence of water vapor in the diver's exhalations. Even a small variation in the specific gravity of the breathing gas can significantly affect the frequency shift.

There is another effect on the diver's speech which must be taken into account. The acoustic impedance of the gas can be derived from Newton's law, by calculating the force δP on a mass-less piston of area A. The mass displaced by the piston is $\rho\ A\ \delta x$, the force $A\ \delta P$, and the acceleration $\delta u/\delta t$ where u is the gas particle velocity, giving

$$\delta P = \rho\ \delta x\ (\delta u/\delta t) = (\rho\ \delta x / A) * (\delta U(x,t) / \delta t) \qquad ...2$$

where U is the volume velocity of the air.

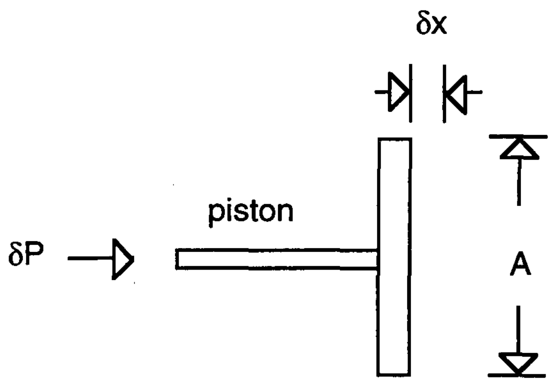

Figure 1 calculation of acoustic impedance

For $U(x,t) = U(x)\, e^{j\omega t}$

$$\delta P / \delta x = j\, \omega\, L\, U \qquad \ldots 3$$

where the acoustic impedance per unit length $j\, \omega\, L = j\, \omega\, \rho\, /\, A$.

So, the acoustic impedance of the gas is proportional to gas density and therefore depth, while the impedance of the vocal tract walls remains constant. Increased coupling at depth causes the vocal tract walls to vibrate which increases the frequency and bandwidth of the formants [2]. This shift is inversely proportional to frequency [3, page 69].

Other communication impediments are also present, such as the divers mask, bubbles and breathing noises. There is a loss of high frequency amplitude due to the energy roll-off of speech (about -10 dB / octave above 1 kHz). The diver suffers a loss of hearing acuity. The diver's pitch is shifted upward, due to coupling of the vocal tract walls, increased vocal intensity (an unconscious response to increased vocal effort and reduced hearing acuity) and psychological factors [4]. The pitch shift is generally not thought to have a significant impact on intelligibility.

3.0 SOLUTION IMPLEMENTED BY CRC

Over the last several decades, a variety of methods have been developed in an attempt to improve the intelligibility of divers speech [5,6]. Central to these methods is the reversal of the nonlinear frequency distortion, which requires deconvolution of the speech. For example, the speech waveform s(n) can be represented as the convolution of the glottal source signal g(n) and the vocal tract impulse response h(n)

$$s(n) = \Sigma_k \; h(n-k) \quad g(k) \qquad \ldots 4$$

The Fourier Transform of the speech signal is the product of these terms

$$S(\omega) = H(\omega) \; G(\omega) \qquad \ldots 5$$

where $S(\omega)$ is the Fourier transform of s(n)
In the process of homomorphic deconvolution, one takes the logarithm of $S(\omega)$.

$$\ln (S(\omega)) = \ln(H(\omega)) + \ln(G(\omega)) \qquad \ldots 6$$

The two terms can then be separated, allowing the term $H(\omega)$ to be independently manipulated. Although this approach provides the necessary means of correcting the speech distortion, it is computationally demanding. In addition, $h(\omega)$ is difficult to interpolate because it is not a well-behaved function. Transformation of $h(\omega)$ can lead to instability in the reconvolution step.

Another method of deconvolving the speech signal involves the use of linear prediction. In this method, an estimate of the speech signal s'(n) is obtained from a weighting function a(k) applied to previous speech samples

$$s'(n) = \Sigma_k \; a(k) \; s(n-k) \qquad \ldots 7$$

The weighting function a(k) is selected to minimize the energy of the error signal e(n) = s(n) - s'(n). Because the weighting function is time dependent, the speech is windowed in short segments (about 20 ms) during which time the speech is assumed to be stationary. The weights are updated once per frame.

e(n) represents the glottal source signal g(n) while the filter a(k) is an estimate of the impulse response h(k). The Fourier Transform of 1/ (1 - a(k)) is an estimate of the spectral envelope of the speech.

A helium speech descrambler can be fashioned by the following sequence of operations.

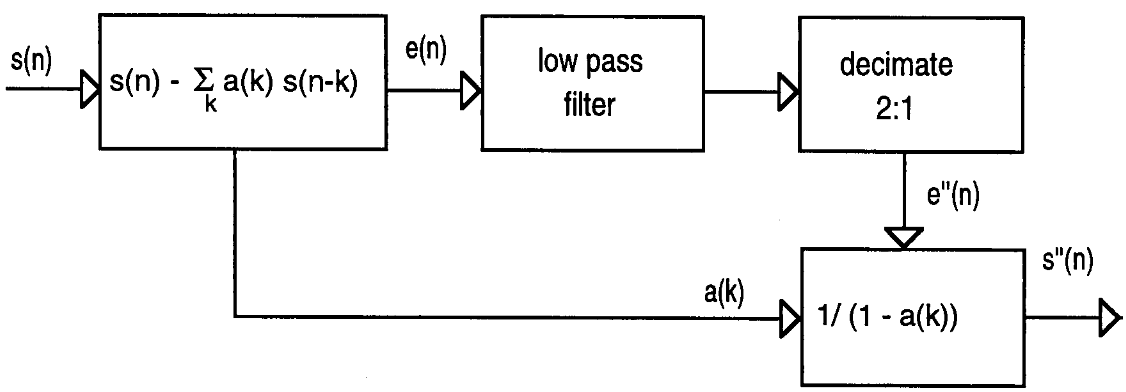

Figure 2 frequency translation by decimation

In this method, a linear translation of two in frequency is obtained by low pass filtering the signal e(n), then decimating by two to produce the signal e"(n). When this signal is applied to the original estimate of the vocal tract frequency response, the result s"(n) contains the original (but low pass filtered) g(n), convolved with h(2n), achieving the desired transformation. Other frequency translations can be obtained by modifying the filter-decimation step.

One limitation of this approach is that the frequency translation is coupled to a proportionate reduction in the output sampling rate. Another limitation is that only linear frequency translations can be obtained. Thirdly, translations other than 2:1 require computationally intensive upsampling and filtering steps. For example, to obtain a 3/4 downshift in frequency, the following sequence of operations is required.

Figure 3 frequency translations of 4:3

The new approach described here achieves frequency translation by indirect manipulation of the filter coefficients. The filter coefficients a(m) may not be transformed directly otherwise instability may result. They are derived from the autocorrelation of the speech signal

$$\phi(m) = \Sigma_n \; s(n) * s(n+m), \quad 0 <= |m| <= L \qquad ...8$$

The filter coefficients are obtained by applying the Leroux-Gueguen algorithm [7] to obtain the lattice filter coefficients k(m), then converting these to the direct form filter coefficients a(m) [8]. The advantage of doing the transformation in this manner is that the coefficients k(m) are bounded by 1 and can thus be conveniently calculated using a fixed point processor, and that the resulting filter is guaranteed to be stable.

$\Phi(e^{j\omega})$ the Fourier Transform of $\phi(m)$, is the power spectral density of the speech signal.

$$\Phi(e^{j\omega}) = \Sigma_m \; w(m) \; \phi(m) \; e^{-j\omega m} \qquad ...9$$

where w(m) = 1 for m = -M to M, 0 elsewhere

By transforming $\phi(m)$ into the frequency domain, one can indirectly manipulate the frequency response of the filter. However consider the following experiment.

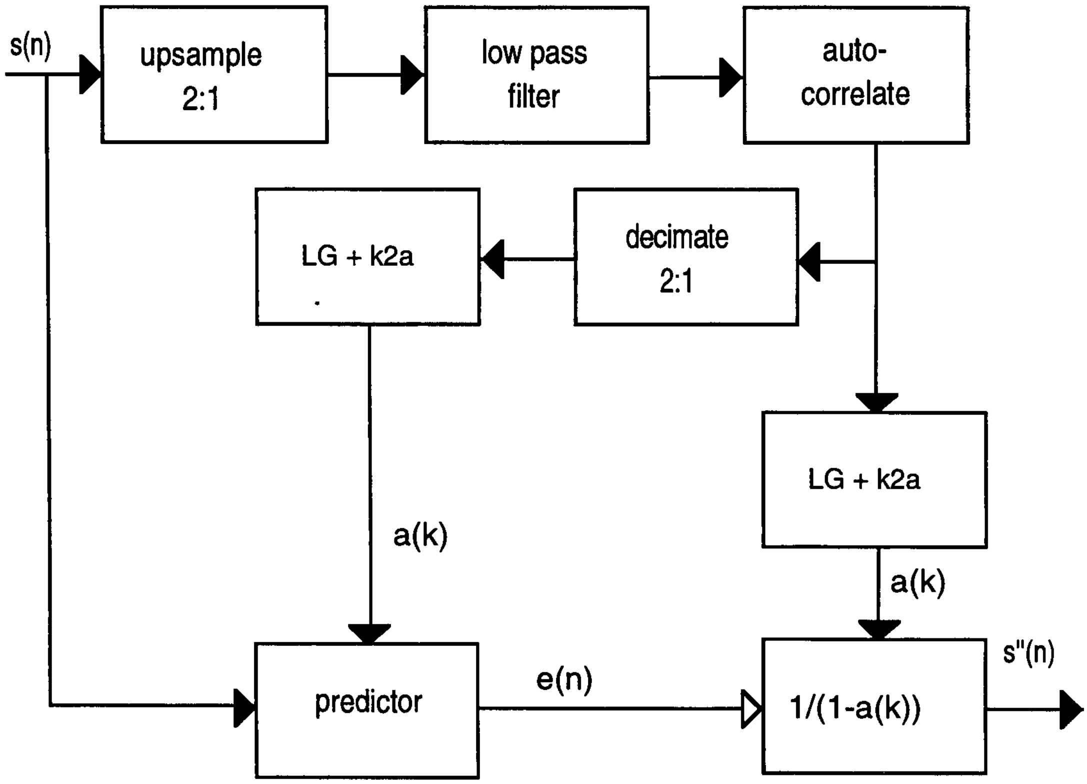

Figure 4 CRC descrambler model 2

In this process, the speech signal is upsampled prior to the calculation of $\phi(m)$. The effect is to divide the index m by two. If one observes the power spectral density of $\phi(m/2)$, one finds that the frequencies are scaled down by a factor of two, as expected from the equation. By downsampling the autocorrelations by two, it is possible to deconvolve the speech signal as in the previous experiment. However, the filter coefficients derived from $\phi(m/2)$ also yield a frequency response which is scaled down by two. By applying e(n) to this filter, the linear frequency distortion is eliminated. This has been done without the necessity of modifying the output sampling frequency.

The power of this approach is revealed by considering the direct relationship between the filter frequency response and scaling of m. For example

$$\Phi'(e^{j\omega}) = F\ (\phi(m/(a+b/(1+\omega/\omega_0)\ )\ )\ ) \qquad \ldots 10$$

where F() represents the Fourier Transform

This transformation will provide a linear frequency shift of 1/a, plus an additional frequency shift of 1/(a+b) for small ω. It can be accomplished by interpolation of $\phi(m)$. Thus, independent compensation can be provided for sound velocity and cavity wall vibration. The resulting filter is guaranteed to be stable. Equivalent manipulation cannot be accomplished by manipulating $\Phi(e^{j\omega})$ because unlike $\phi(m)$, it is not a smooth function and is not easy to interpolate.

Noise in the input signal s(n) will be manifested as noise in the error signal e(n), and will be little affected by the deconvolution operation. Compensation for noise could be accomplished by limiting of the signal e(n) prior to reconvolution.

4.0 DESCRAMBLER DESCRIPTION

The Model 1 descrambler developed by CRC was based on fig 2 and implemented on the TMS320-C25. The computational load limit of this DSP limited the input sampling frequency to 14 KHz and the synthesis filter order to 10. Frequency translations different from 1.5:1 to 2.5:1 were obtained by varying the output sampling frequency slightly, which required linear interpolation of the residual e"(n). As a result, the output bandwidth at the maximum frequency compression was only 2.8 KHz. In addition, interpolation of the residual introduces noise.

The Model 2 descrambler is based on fig 4. In this unit, the input and output sampling frequencies are 16 KHz, and a 20th order synthesis filter is used. Arbitrary frequency compression can be obtained by interpolation of the autocorrelation coefficients (see Appendix 1). The simulations implement linear frequency translations from 1.6:1 to 2.1:1 in steps of 0.1. The hardware implements linear frequency translations from 1.6:1 to 2.4:1 in steps of 0.1, and includes a 1:1 setting.

Two versions of the model 2 descrambler were developed, a fixed point version using the TMS320-C50 DSP and a floating point version using the TMS320-C31 DSP. All schematics and a cassette containing all of the software developed for the project are attached. One unit of each version has been previously delivered to DCIEM.

Both units are equipped with manual rotary switches used for "depth" settings, although no attempt was made to calibrate the switches. The "best" settings for each depth found during the subjective tests are recorded in the attached subjective test report.

Both versions of the descrambler use the same A/D converter (although the AE2000 is the newer part) with similar analog circuits. Both have 32k bytes of ROM and 32k bytes of RAM. They are the same circuits which are used in other projects at CRC as DSP engines for processing acoustic signals.

The reasons for moving to a floating point processor are accuracy and ease of development. Fixed point calculations create scaling problems and induce round-off error (and thereby add noise). A floating point processor eliminates the need for hand crafting of the software and allows development in a high level language.

The processing power limitation of both the fixed and floating point processor required elimination of the upsampling step. The upsample/low pass/autocorrelate steps in fig 4 are equivalent to upsampling the input speech to 32k samples/sec (although the effective bandwidth is still less than 8 kHz), and are thus very demanding in real-time. This part of the algorithm makes interpolation of the autocorrelation coefficients easier, and improves the accuracy of the vocal tract filter. Autocorrelations were performed instead on the 16k samples/sec signal.

5.0 ANALYSIS AND TEST RESULTS

A FFT of the autocorrelations produces the speech power spectrum, of which fig 5 is an example (fig 5 also shows the effect of truncating the autocorrelations, since the speech power spectrum is by definition a positive function).

Fig 5 is very revealing, in that it shows virtually no trace at all of the second formant of the speech signal. This is due to several factors, (a) the previously mentioned drop in speech energy above 1 kHz, (b) the microphone frequency response in helium and audio frequency response of the diver communication system and (c) , suppression of the higher formants by the face mask (which makes the speech sound muffled). Suppression of the second and higher formants will make the descrambler much more sensitive to wide band background noise. Appendix 2 shows the frequency response of the M-101/AIC microphone in air.

Autocorrelations derived from the 32 k sample/sec signal produce a "smooth" function of which fig 6 provides a few examples. Using autocorrelations derived directly from the 16 k sample/sec signal complicates the problem of interpolation. Fig 6 also shows the effect of the low pass filter on the accuracy of the calculation.

Linear interpolation of the autocorrelations derived from the 16k samples/sec signal will often produce values less than the correct values, which will tend to reduce the Q of the formants in the synthesis filter and distort their frequencies slightly. In recognition of this, cubic interpolation of the autocorrelations (Appendix 3) was tested. No noticeable improvement of the intelligibility was observed when using the speech samples recorded at DCIEM. Further, increasing the order of the synthesis filter beyond 20 did not yield any noticeable improvement.

Nevertheless, increasing the sampling frequency and bandwidth of the input speech is thought to be an important step to improving the performance of the descrambler.

No attempt was made to implement the nonlinear frequency translation algorithm described in section 3. Instead, only linear frequency translations were used. This was done for two reasons, (a) because there was no accurate model on which to base the nonlinear translation, and (b) because the degradations in the descrambler were dominated by other factors, particularly face mask distortion, rendering it very difficult to gauge the improvement resulting from empirical attempts to remove nonlinear frequency distortion.

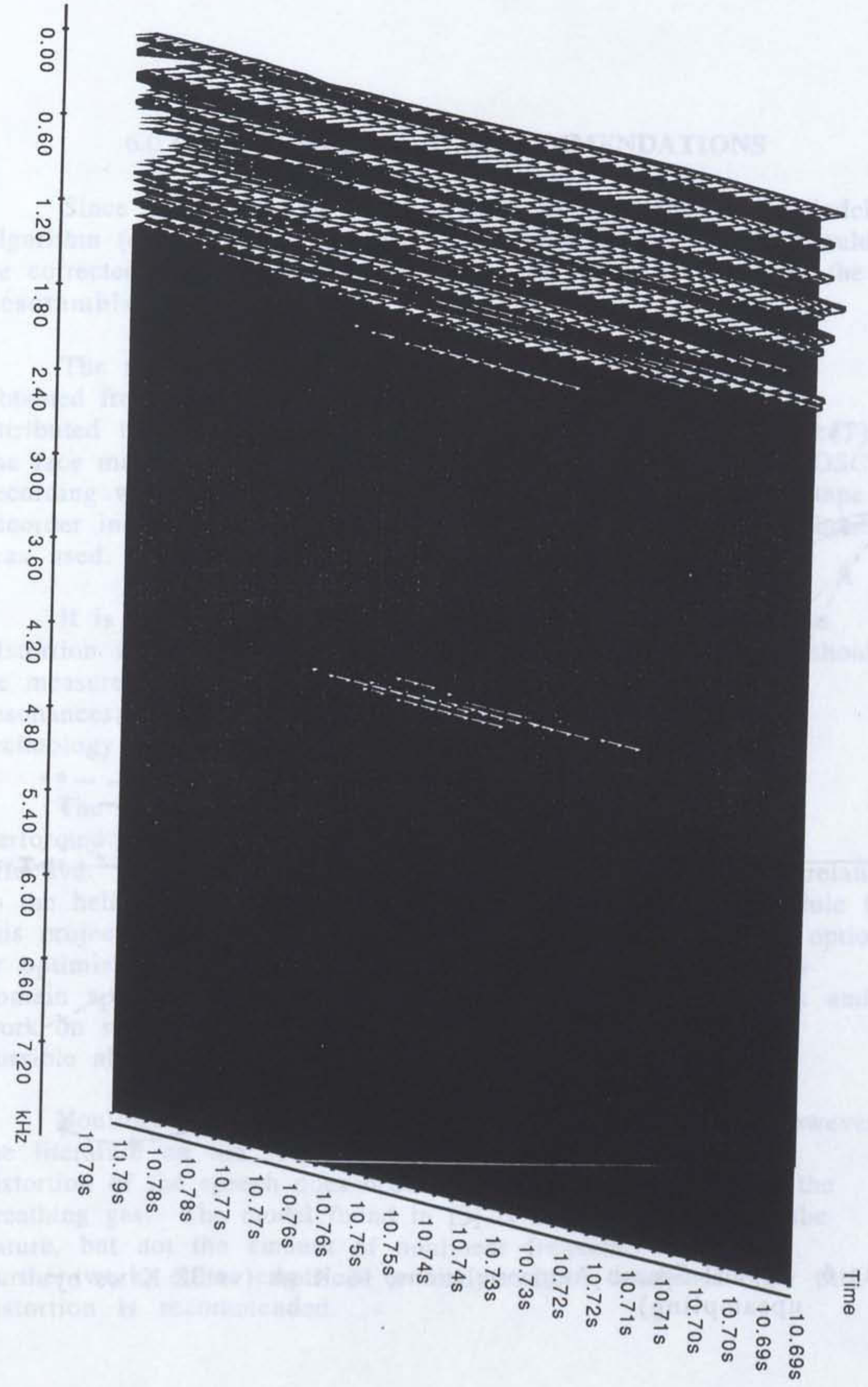

Fig 5 Speech Spectrogram for vowel "ee" in feet

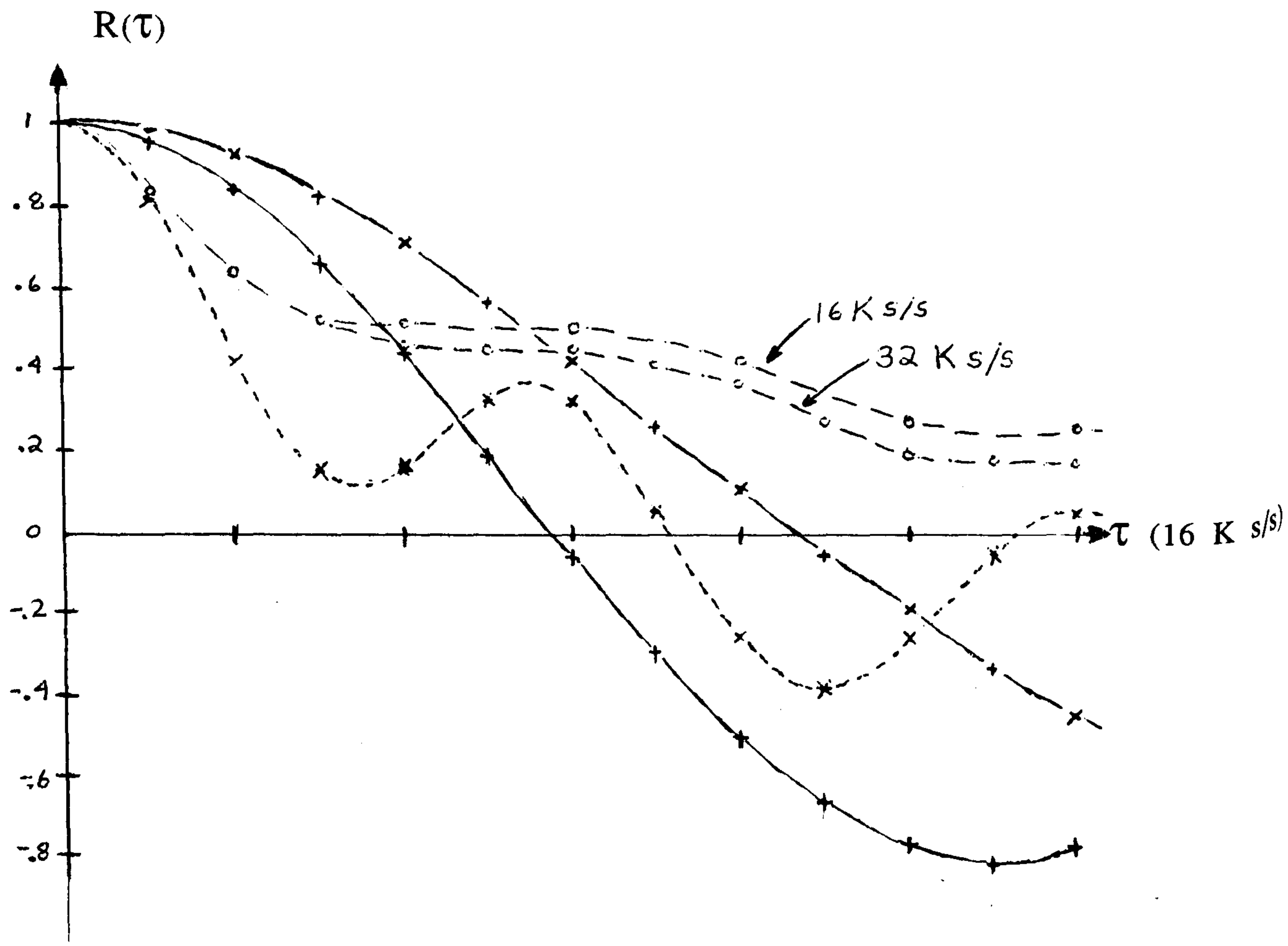

Fig 6 Typical Speech Autocorrelations, 16 K s/s (vs 32 K s/s by upsampling)

6.0 CONCLUSIONS AND RECOMMENDATIONS

Since the completion of the project, an oversight in the Model 2 algorithm (described in Appendix 4) has been noticed. This should be corrected to see if it has any bearing on the performance of the descrambler.

The major source of the difference between the results obtained from the NOSC recording and the DCIEM recordings is attributed to the diver equipment, in particular distortion caused by the face mask and microphone used by the Navy divers. The NOSC recording was made in a chamber without face masks using a tape recorder in the chamber. It is not known what type of microphone was used.

It is recommended that as a first step to further work, the distortion introduced by the Navy face mask and microphone should be measured. Modifications to the face mask which reduce resonances in the face mask cavity and alternate microphone technology should be considered.

The results indicate that the CRC Model 2 and Hele unit performed at about the same level and that neither was very effective. In fact, both units actually decreased intelligibility relative to the helium speech itself. The relatively tight delivery schedule for this project did not allow time for exploring other algorithmic options or optimizing the current algorithm. In particular, a frequency domain approach to descrambling offers intriguing possibilities and work on such algorithms is recommended. A description of a possible algorithm is given in Appendix 4.

Nonlinear frequency translations were not explored. However the literature on the subject clearly indicates that nonlinear distortion of the speech does occur due to the high density of the breathing gas. The model found in [3] is sufficient to predict the nature, but not the amount of nonlinear frequency distortion. Further work, either empirical or analytical, to compensate for this distortion is recommended.

7.0 REFERENCES

1. B. Gilliam, "Soulmates in the Abyss", Scuba Times, Sept/Oct 1993.

2. H.B. Rothman et al, "Speech Intelligibility at high helium-oxygen pressures", Undersea Biomedical Research, Vol. 7, No. 4, pg. 265, Dec. 1980.

3. J.L. Flanagan, Speech Analysis, Synthesis and Perception, Springer Verlag, ISBN 3-540-05561-4, 1972.

4. H. Hollien et al, "Voice fundamental frequency levels of divers in helium-oxygen speaking environments", Undersea Biomedical Research, Vol. 4, No. 2, pg. 200, June 1977.

5. S. Seneff, "System to Independently Modify Excitation and/or Spectrum of Speech Waveform Without Explicit Pitch Extraction", IEEE Trans. ASSP, Vol. 30, No. 4, pg. 566, Aug. 1982.

6. M. Richards, "A System for Helium Speech Enhancement Using the Short-Time Fourier Transform", Proc. ICASSP'81, Vol. 3, pg. 1097, April 1981.

7. J. Leroux, C. Gueguen, "A Fixed Point Computation of Partial Correlation Coefficients", IEEE Trans ASSP, pg. 257, June 1977.

8. L. Rabiner, R. Schafer, Digital Processing of Speech Signals, Prentice-Hall, ISBN 0-13-213603-1, 1978, pg. 443.

APPENDIX 1, NON-INTEGER FREQUENCY SHIFTS OBTAINED BY LINEAR INTERPOLATION OF THE AUTOCORRELATION FUNCTION.

A frequency down shift of 2:1 in the synthesis filter is obtained by simple virtue of the fact that the autocorrelation function is calculated on a speech signal which has been upsampled to 32 k samples/sec, whereas the clock rate of the synthesis filter is 16 k samples/sec. In order to increase the frequency compression above 2:1 the autocorrelation function can be interpolated to obtain values at intervals less than one sampling interval T. To obtain compression ratios below 2:1, intervals greater than one sampling interval T are used. For example, a compression ratio near 1.7 (2 * 16 / 19) can be obtained by the following interpolation. The weights have been expressed so that precision is not lost if the calculation is done in fixed point. The numbers in brackets (X) indicate the autocorrelation at lag X.

interpolated values		weighted sum of original autocorrelations
(0)	=	(0)
(1)	=	13/16 (1) + 3/16 (2)
(2)	=	5/8 (2) + 3/8 (3)
(3)	=	7/16 (3) + 9/16 (4)
(4)	=	1/4 (4) + 3/4 (5)
(5)	=	1/16 (5) + 15/16 (6)
(6)	=	7/8 (7) + 1/8 (8)
(7)	=	11/16 (8) + 5/16 (9)
(8)	=	1/2 (9) + 1/2 (10)
(9)	=	5/16 (10) + 11/16 (11)
(10)	=	1/8 (11) + 7/8 (12)
(11)	=	15/16 (13) + 1/16 (14)
(12)	=	3/4 (14) + 1/4 (15)
(13)	=	9/16 (15) + 7/16 (16)
(14)	=	3/8 (16) + 5/8 (17)
(15)	=	3/16 (17) + 13/16 (18)
(16)	=	(19)

Table I, Interpolation of Autocorrelations for 1.7:1 Frequency Compression

APPENDIX 2, DISCUSSION OF M-101/AIC MICROPHONE AND AUDIO SYSTEM FREQUENCY RESPONSE FOR DIVER COMMUNICATIONS

The M-101/AIC microphone frequency response was swept in an air medium (facilities do not exist at CRC to do this experiment in heliox). If the response in helium is representative of that in air, the following is its frequency response, rising at about 10 dB/octave up to perhaps 2.5 KHz, then falling.

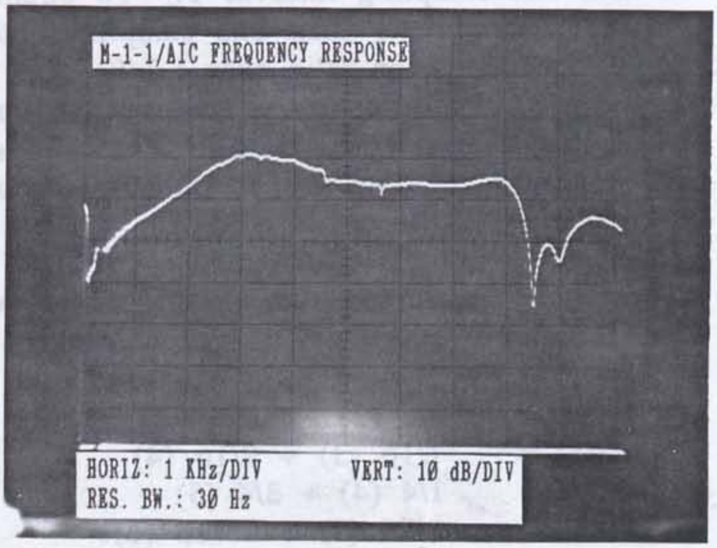

Figure 7 M-101/AIC Microphone Frequency Response

The vocal tract formants which give rise to the various speech sounds (including the consonants) are the result of the vocal tract acting like an organ pipe, with a node at the glottis and an antinode at the lips. To illustrate:

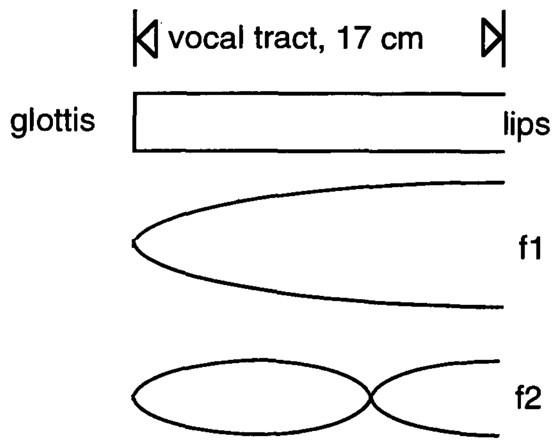

Figure 8 Origin of Formants in Speech Signal

For a typical male vocal tract length of 17 cm and velocity of sound in air of $3.5 * 10^4$ cm/sec the first formant frequency will be:

$$f1 = 3.5 * 10^4 / 17 / 4 = 515 \text{ Hz.} \qquad ...11$$

The succeeding formants will be at odd multiples of the first, or 1545 Hz, 2575 Hz and so on. Formant frequencies at these values will produce the unstressed vowel such as the "e" in the casually spoken phrase "the book". The movement of the jaws, lips and tongue pushes the formants from the nominal values. The second formant can range up to at least 2800 Hz in a male speaker for the vowel "e" in the word "helium". In heliox, these frequencies will be roughly doubled.

It is generally accepted that to obtain reasonable intelligibility in speech, the first three formants must be preserved. This means that the entire audio system used for diver communication must have a 3 dB bandwidth of at least 8000 Hz. Judging by the microphone frequency response, this requirement is not being met.

APPENDIX 3, CUBIC INTERPOLATION OF THE AUTOCORRELATION FUNCTION

Linear interpolation of the autocorrelation function will generally produce an underestimate of the correct value. In order to improve the interpolation a curve fitting formula can be used. For example, Newton's forward interpolation formula for a cubic polynomial is expressed as:

$$y = Rx + \mu\,\Delta Rx + \mu\,(\mu - 1)\,\Delta^2 Rx\,/\,2 + \mu\,(\mu - 1)\,(\mu - 2)\,\Delta^3 Rx\,/\,6 \qquad \ldots 12$$

Now, given four observations of autocorrelations Rx, Rx+1, Rx+2, Rx+3, find an intermediate value between Rx+1 and Rx+2:

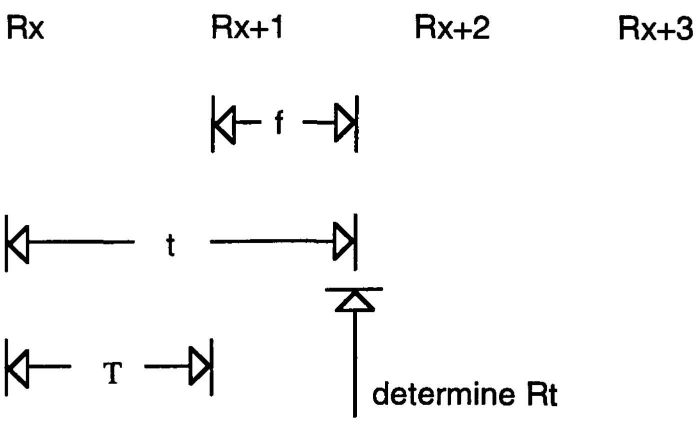

Fig 9, Cubic Interpolation of Autocorrelations

Now, $\mu = t / T = f + 1$, where f is a fraction between 0 and 1, so y(t) can be expressed as

$$y = Rx + (f + 1)\,\Delta Rx + f\,(f + 1)\,\Delta^2 Rx\,/\,2 + f\,(f + 1)\,(f - 1)\,\Delta^3 Rx\,/\,6 \quad \ldots 12a$$

The constants involving f can be precalculated for the particular frequency shift desired. To obtain the differences, use a matrix as follows:

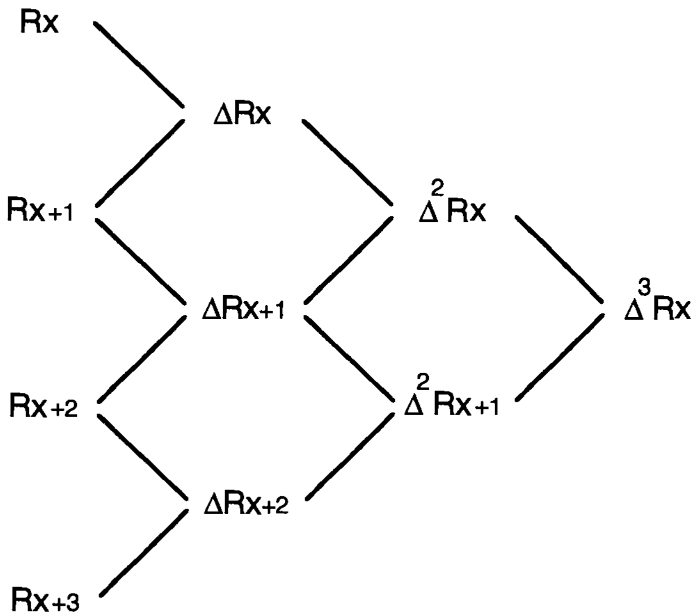

Fig 10, Partial Differences of the Autocorrelation Function

To calculate interpolated values between R_0 and R_1, use the property that $R_{-1} = R_1$.

APPENDIX 4, FREQUENCY DOMAIN APPROACH TO FREQUENCY SCALING

No matter what approach is used for frequency scaling, the speech must be deconvolved as much as possible to separate the pitch signal from the formant information. This is the case because the pitch of the voice is a function of the vocal chord mass and surrounding muscles, and is not affected by the helium in the same way or to the same degree as the formant frequencies.

The linear prediction technique accomplishes deconvolution because only short term correlations are used. Viewed in the frequency domain, it succeeds because the resonances of the vocal tract tend to occur at frequencies significantly higher than the pitch frequency. This assumption is not entirely accurate, although the frequency shift caused by the helium works in its favor.

In addition, the windowing of the speech signal causes the pitch to interfere with the calculation (depending on exactly where the window falls relative to the pitch period). To compound the problem, deconvolution using autocorrelations is only theoretically correct if the speech is stationary, which makes the prediction work rather poorly during rapid transitions.

The result is that the deconvolution is imperfect. Any vocal tract information which is left in the residual will not be corrected for frequency shift. Further, any perturbations in the vocal tract filter caused by the pitch will be frequency translated. Finally, the computation of the vocal tract filter must be done in such a way that the synthesis filter is stable.

Equation 5 provides an alternative approach to synthesis. Once the speech has been deconvolved, frequency scaling of the vocal tract spectrum (obtained by equation 9) can be accomplished directly in the frequency domain. The residual signal is translated by FFT to the frequency domain and then multiplied by the modified vocal tract spectrum to synthesize the descrambled speech.

The hook (there is always one) is that discontinuities will occur at frame boundaries due to the insults that have been perpetrated on the speech spectrum. However, these can be effectively eliminated using a weighted overlap and add technique, as shown below.

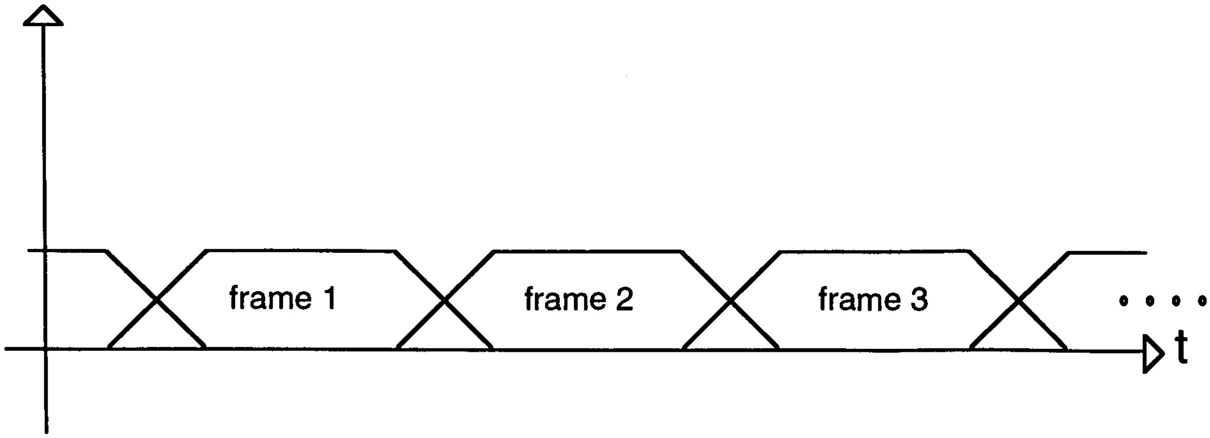

Fig 11, Frame Overlap and Add Technique

Remnants of vocal tract information in the residual cause the residual spectrum to have peaks and valleys (it should theoretically be flat if the deconvolution is perfect). Nonlinear compression of the residual signal will help to remove some of them (and may also help deal with background noise). In addition, the residual signal should be low pass filtered prior to reconvolution. (This is an oversight in fig 4 which although small, may have affected the performance of the CRC descrambler. There should be a low pass filter between the predictor and the synthesizer). Because the bandwidth of the residual is wider than that of the vocal tract filter, out-of-band noise will be synthesized in the upper half of the frequency spectrum. The low pass filtering is easily accomplished in the frequency domain by truncating the residual.

DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when the overall document is classified)

1. ORIGINATOR (the name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g. Establishment sponsoring a contractor's report, or tasking agency, are entered in section 8.)	2. SECURITY CLASSIFICATION (overall security classification of the document, including special warning terms if applicable)
COMMUNICATIONS RESEARCH CENTRE 3701 Carling Avenue, P.O. Box 11490, Station H OTTAWA, Ontario K2H 8S2	UNCLASSIFIED

3. TITLE (the complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S,C or U) in parentheses after the title.)

HELIUM SPEECH DESCRAMBLER FINAL REPORT (U)

4. AUTHORS (Last name, first name, middle initial)

BRYDEN, KAREN R., HASSANEIN, HISHAM., MARCEAU, FRANÇOIS

5. DATE OF PUBLICATION (month and year of publication of document)	6a. NO. OF PAGES (total containing information. Include Annexes, Appendices, etc.)	6b. NO. OF REFS (total cited in document)
JUNE, 1995	29	

7. DESCRIPTIVE NOTES (the category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

DND FINAL REPORT (CRC REPORT 95-004)

8. SPONSORING ACTIVITY (the name of the department project office or laboratory sponsoring the research and development. Include the address.)

DEFENCE RESEARCH ESTABLISHMENT OTTAWA
DEPARTMENT OF NATIONAL DEFENCE OTTAWA
OTTAWA, ONTARIO K1A 0Z4

9a. PROJECT OR GRANT NO. (if appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant)	9b. CONTRACT NO. (if appropriate, the applicable number under which the document was written)
041LA	

10a. ORIGINATOR'S DOCUMENT NUMBER (the official document number by which the document is identified by the originating activity. This number must be unique to this document.)	10b. OTHER DOCUMENT NOS. (Any other numbers which may be assigned this document either by the originator or by the sponsor)
CRC REPORT No 95-004	

11. DOCUMENT AVAILABILITY (any limitations on further dissemination of the document, other than those imposed by security classification)

(X) Unlimited distribution
( ) Distribution limited to defence departments and defence contractors; further distribution only as approved
( ) Distribution limited to defence departments and Canadian defence contractors; further distribution only as approved
( ) Distribution limited to government departments and agencies; further distribution only as approved
( ) Distribution limited to defence departments; further distribution only as approved
( ) Other (please specify):

12. DOCUMENT ANNOUNCEMENT (any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). However, where further distribution (beyond the audience specified in 11) is possible, a wider announcement audience may be selected.)

N/A

13. **ABSTRACT** (a brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), or (U). It is not necessary to include here abstracts in both offical languages unless the text is bilingual).

During diving operations, Canadian Navy divers are kept in constant contact with the surface tender via communication lines bundled with the breathing air hoses. Under certain circumstances where depth prevents the use of compressed air as a breathing gas, a mixture of helium and oxygen is used, consequently rendering the divers' speech unintelligible. A helium speech "descrambler" is used to remove as much of the distortion as possible.

A helium speech descrambler made by Hele Co. is used by the Canadian Navy. This unit is expensive, and is often found to operate ineffectively. Other available descramblers such as those made by Stocktronics are extremely expensive. In order to provide a low cost alternative to these descramblers, the Defence and Civil Institute of Environmental Medicine (DCIEM) tasked the Communications Research Centre (CRC) to investigate the problem and design a new descrambler.

CRC had previously simulated a new descrambler and used it successfully to descramble a sample of helium speech recorded at the Naval Ocean System Centre (NOSC) in San Diego Ca. This algorithm was transferred to a digital signal processing (DSP) chip and tested during a set of helium dives performed at the Fleet Diving Unit (Atlantic) in Shearwater NS in Sept. 1993. The Model 1 unit performed as well as, and in some cases better than the Hele unit, in spite of its simplicity.

Recordings made during the trials were brought back to Ottawa and used during the development of a Model 2 descrambler. In Jan. 1994, during subsequent helium chamber dives at DCIEM, divers were asked to speak a set of isolated words from a subjective listening test at representative depths. The divers speech was recorded and later passed through both the Hele descrambler and the Model 2. A comparison of the two descramblers was obtained by playing back the descrambled speech to test subjects who tried to identify each word among a set of similar sounding words. This report details the results of the subjective tests and the algorithms, hardware and software developed by CRC.

14. **KEYWORDS, DESCRIPTORS or IDENTIFIERS** (technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus. e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus-identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

Diver Communication

Helium Speech

Mixed Gas Diving