Adaptive Speech Enhancement using Diverse Processing in Non-linearly Distributed Sub-bands

Amir Hussain, Douglas R.Campbell and T.J.Moir
Department of Electronic Engineering and Physics,
University of Paisley, High St., Paisley PA1 2BE, Scotland U.K.
Corresponding author's email: amir@diana22.paisley.ac.uk

ABSTRACT

A multi-microphone sub-band adaptive speech enhancement scheme using a human cochlear model is presented. The effect of distributing the sub-bands non-linearly as in humans is investigated. A new robust metric is developed in order to automatically select the best form of diverse processing within each sub-band. Comparative results achieved in simulation experiments demonstrate that the proposed scheme employing diverse processing in cochlear spaced sub-bands is capable of significantly outperforming conventional noise cancellation schemes.

1. INTRODUCTION

Background noise contamination of speech signals reduces the Signal to Noise Ratio (SNR) of for example, hands-free telephones, portable phones, and security screens. Speech recognition systems in particular, are known to experience problems due to levels of background noise that are quite acceptable to human listeners [6]. In addition to the noise level, the presence of multiple noise sources, reverberant environments, moving noise sources and statistically non-stationary noise sources considerably complicates the situation.

Classical speech enhancement methods based on full-band multi-microphone noise cancellation implementations which attempt to model acoustic path transfer functions can produce excellent results in anechoic environments with localized sound radiators [5], however performance deteriorates in reverberant environments. Multi-band processing has been found to be important in combating reverberation effects [9] [10]. Adaption is necessary to compensate for changing noise fields [2] [9] due for example to, non-Gaussian sources, source/sensor motion, or time-varying acoustic paths. Multi-sensor methods are necessary to compensate for reverberation [12] and speech/noise spectral overlap [2].

Studies of noise in automobiles [8][9][10] have found that the noise correlation between two microphone locations was high for frequencies below 500Hz, and decreased gradually with virtually no correlation above 2kHz. Wallace and Goubran [11] applied the classical full-band linear FIR based correlated noise canceller, adapted using the Least Mean Square (LMS) algorithm, to recorded automobile noise and obtained significant noise reduction in the low frequency range but at high frequencies, where both the correlation and noise energy were low, the noise increased. In some situations, as reflected in the measurements from office and automobile environments recently reported in [2], significant short term correlation may also occur at high frequencies depending upon the relative locations of the microphones and the nature of the noise sources. The above evidence implies that processing appropriate in one sub-band, may not be so in another. This therefore supports a general approach involving the use of diverse processing in frequency bands dependent on the correlation between the in-band signals from multiple sensors. Dabis et al [5] used closely spaced microphones in a full-band adaptive noise cancellation scheme involving the identification of a differential acoustic path transfer function during a noise only period in intermittent speech. A Multi-Microphone Sub-Band Adaptive (MMSBA) speech enhancement system has been described which extends this method by applying it within a set of sub-bands provided by a filterbank [2][7][10]. Even non-optimised MMSBA processing has shown the potential to yield more than 6dB SNR improvements over conventional full-band methods in real reverberant environments.

In this paper, the MMSBA system has been further developed by employing diverse Sub-Band Processing (SBP) in order to allow inter-channel features within the sub-bands to influence the subsequent processing. In order to realize this, a robust practical metric has been developed, which is capable of real-time implementation, based on the inter-channel Magnitude Squared Coherence (MSC) relationship in order to automatically select the best SBP option. The choice of SBP options include: (i) no processing; (ii) intermittent coherent noise canceller; and (iii) continuous incoherent noise canceller. The effect of spacing the sub-bands non-linearly as in humans, according to a published cochlear function is also investigated.

2. THE PROPOSED MMSBA SCHEME BASED ON COCHLEAR MODEL EMPLOYING DIVERSE SBP

Two or more relatively closely spaced microphones may be used in an adaptive noise cancellation scheme [5][10]

to identify a differential acoustic path transfer function during a noise only period in intermittent speech. The MMSBA speech enhancement system applies the method within a set of sub-bands provided by a filter bank as shown in Figure 1. The filter bank can be implemented using various orthogonal transforms or by a parallel filter bank approach. The sub-bands can be distributed in a linear or a non-linear fashion.

It is assumed in this work that the speaker is close enough to the microphones so that room acoustic effects on the speech are insignificant, that the noise signal at the microphones may be represented as a point source modified by two different acoustic path transfer functions H_1 and H_2, and that an effective voice activity detector (VAD) is available.

2.1 Cochlear Modeling

In previous work [2] [3] [7] [10], the sub-band filters were spaced linearly in the frequency domain. The human cochlea, which evolved to deal with all sounds available to the human ear [4], has been modeled by Ghitza [17] who proposed use of the logarithm function for approximating the cochlear distribution of filters. Greenwood [18] has presented the following more accurate function for the spacing of the filters in the *mammalian* cochlea:

$$F(x) = A(10^{ax} - k) \text{ Hz}$$

where x is the proportional distance from 0 to 1 along the cochlear membrane, A, a and k are constants based on empirical knowledge of the cochlea, and $F(x)$ are the upper and lower cut-off frequencies for each filter obtained by the limiting value of x. For the human cochlea, values of A=165.4, a=2.1 and k=0.88 are used, and this is confirmed by Allen [19]. In this work, the sub-bands are achieved by modifying the spectra of the FFT (or DCT) of the input signals, and the number of filters is therefore limited by the size of the FFT.

2.2 Diverse SBP Options

The sub-band processing (SBP) can be accomplished in a number of ways, for example:

1. *No Processing*: Examine the noise power in a sub-band and if below (or the SNR above) some arbitrary threshold, then the signal in that band need not be modified.

2. *Intermittent coherent noise canceller*: If the noise power is significant and the noise between the two channels is significantly correlated in a sub-band, then perform adaptive intermittent noise cancellation, wherein an adaptive filter may be determined which models the differential acoustic-path transfer function between the microphones during the noise alone period. This can then be used in a noise cancellation format during the speech plus noise period to process the noisy speech signal. This scheme illustrated in Figure 1 can be described mathematically as follows. Assuming N, S, P, R represent the z-transforms of the noise signal, speech signal, primary signal and reference signal, respectively. The primary and reference signals in each sub-band are thus

$$P = B(S + H_1 N) \quad ; \; R = B(S + H_2 N)$$

The transformed error signal is thus,

$$E = B[(1 - H_3)S + (H_1 - H_3 H_2)N]$$

which is a frequency domain error, weighted by the band-limiting transfer function B, and H_3 represents the sub-band adaptive filter. The Mean Squared Error (MSE) function is,

$$J_E = (2\pi j)^{-1} \oint_{|z|=1} E.E^* z^{-1} dz$$

The sub-band noise cancellation problem is thus, to find an H_3 such that within the sub-band defined by B, the variance of J_E is minimised. During a noise only period $S = 0$, defining the noise spectral density Φ_{nn}, then

$$J_E = (2\pi j)^{-1} \oint_{|z|=1} B(H_1 - H_3 H_2)\Phi_{nn}(H_1 - H_3 H_2)^* B^* z^{-1} dz$$

which is minimised in the least squares sense when

$$H_3 = (BH_1)(BH_2)^{-1}$$

That is, H_3 is a band-limited transfer function that minimises the noise power in E. Now using H_3 as a fixed processing filter when speech and noise are present ideally gives:

$$E = B(1 - H_3)S$$

where the output E is a noise reduced, filtered version of the sub-band speech signal. This approach will fail if $H_1 = H_2$, however in practical situations such acoustic path balancing is difficult to achieve.

3. *Incoherent noise canceller*: If the noise power is significant but not highly correlated between the two channels in a sub-band, then an incoherent noise cancellation approach [14][16] may be applied during the noisy speech period. Since in this case, the primary channel noise component BH_1N is uncorrelated with the reference channel noise component BH_2N, the filtered reference is an estimate of the sub-band speech signal S.

In this paper, we incorporate the above three SBP options and implement the processing using the Least Mean Squares (LMS) algorithm [15] to perform the adaption.

2.3 Metric for Selecting SBP based on Magnitude Squared Coherence (MSC)

The Magnitude Squared Coherence (MSC) has been used by Allen et al [13] to correct the magnitude of a reverberant signal. Recently, Bouquin and Faucon [1] have applied the MSC for noise reduction and successfully employed it as a VAD for the case of spatially uncorrelated noises. In this work, we propose

the use of a modified MSC as a part of a system for selecting the best SBP option in a MMSBA speech enhancement system.

Assuming that the speech and noise signals are independent, the observations received by the two microphones, as shown in Figure 1, may be written as:

At mic 1: $x_1 = s_1 + n_1$; and, at mic 2: $x_2 = s_2 + n_2$

where s_i and n_i (i=1,2) represent the clean speech signal and the disturbing additive noise, respectively.

For each block l and frequency bin fk , the coherence function is given by:

$$\rho(fk,l) = \frac{S_{x_1x_2}(fk,l)}{\sqrt{S_{x_1x_1}(fk,l)S_{x_2x_2}(fk,l)}}$$

where $S_{x_1x_2}(fk,l)$ is the cross-spectral density, $S_{x_1x_1}(fk,l)$ and $S_{x_2x_2}(fk,l)$ are the auto-spectral densities; which can be estimated using a simple recursive calculation on a block by block basis:

$$S_{x_ix_j}(fk,l) = \beta S_{x_ix_j}(fk,l-1) + (1-\beta)X_i(fk,l)X_j^*(fk,l), \quad i,j = 1,2$$

where β is a forgetting factor. During the noise alone or speech free period, for each overlapped and Hann windowed block l we compute the Magnitude Squared Coherence (MSC) at each of the frequency bins fk . $k = 0,\ldots,L/2$, (where L=256 corresponds to the length of the short term FFT) as:

$$\mathrm{MSC}(fk,l) = \frac{|S_{x_1x_2}(fk,l)|^2}{S_{x_1x_1}(fk,l)S_{x_2x_2}(fk,l)}$$

which is then averaged over all the previous overlapped blocks to give (at each frequency bin):

$$\overline{\mathrm{MSC}}(fk) = \frac{1}{l}\sum_{i=1}^{l}\mathrm{MSC}(fk,i)$$

The above *adaptively averaged* MSC criterion can thus be used as a means for determining the level of correlation between the disturbing noises in various frequency bands (by averaging the above MSC over each respective sub-band), during the noise alone period in intermittent speech. The subsequent form of processing in each respective frequency band can therefore be selected between the intermittent coherent noise canceller and the incoherent noise canceller SBP options, on the basis of the adaptive inter-channel correlation measure.

On initial trials, a threshold value of 0.6 for the adaptive MSC has been found to be suitable for distinguishing between highly correlated and weakly correlated sub-band noise signals. For 50% block overlap, a forgetting factor of $\beta = 0.8$ has been found to be adequate, which compares well with the figures reported by Bouquin [1].

3. SIMULATION RESULTS

Simulated Room Reverberant Data:

The impulse responses between the noise source and the two microphones were calculated by an image program [2] which simulates room acoustics using room dimensions, reflection coefficients and source/receiver locations as parameters. At a sampling rate of 10kHz, realistic room responses would be of length > 1024, but for testing purposes a length of 256 was selected. The room was modelled as a (6x5x4)m rectangular enclosure. The walls, floor and ceiling were given the same reflection coefficient value of 0.6 to generate a medium room reverberation level ($T_{60} \cong 0.35$s), and the noise-to-microphones (NTM) and microphone-to-microphone (MTM) spacing were set to 1m and 15cm respectively. Two microphone signals were then generated by convolving a white noise sequence with each of the simulated impulse responses to yield the primary and reference noise signals, which were then added to an anechoic speech signal. The initial SNR was fixed at -3dB, and a noise alone period was manually labelled comprising the first 1024 samples. For this particular test case with the selected MTM, NTM, and noise orientation angle values, the MSC relationship between the two microphone signals during the noise-alone period was found to exhibit a significant level of correlation (>0.6) across all the frequency bands as shown in Figure 2. Three noise cancellation systems were compared namely:

1. The conventional full-band noise canceller intermittently adapted using the LMS algorithm (FBLMS). The order of the adaptive filter was chosen to be 256.
2. A two sensor MMSBA system with four linearly spaced sub-bands employed intermittent adaptive LMS update in each sub-band (termed Linear Multi-Band LMS (LMBLMS)) in order to effectively cancel the correlated noise in those sub-bands. The sub-band filter order was set to 256/4.
3. A cochlear model based MMSBA system with four cochlearly spaced sub-bands employing an intermittent adaptive LMS update in each sub-band (termed Cochlear Multi-Band LMS (CMBLMS)).

System	FBLMS	LMBLMS	CMBLMS
SNR improv.	1.8dB	7.6dB	10.2dB

Table 1: Performance Comparison of adaptive noise cancellers for "realistic" room data.

As can be seen from Table 1, for this test case the use of a cochlear based CMBLMS system gives the best performance in cancelling the room reverberant noise compared to the LMBLMS and the conventional

FBLMS noise cancellers. Informal listening tests also showed the CMBLMS processed speech to be both enhanced in SNR and of better perceived quality than that obtained by the other methods.

4. CONCLUSIONS

A multi-microphone sub-band adaptive (MMSBA) speech enhancement system based on the human cochlear model has been presented. An adaptively estimated inter-channel MSC measure has been proposed for selecting the best form of processing within each cochlearly distributed sub-band. Comparative results achieved in simulation experiments demonstrate that the proposed MMSBA scheme employing diverse processing in cochlear spaced sub-bands is capable of significantly improving the output SNR of speech signals with no additional distortion apparent, compared to conventional noise cancellation schemes. Current experiments are using speech recognizers and real human subjects to further assess and formally quantify the intelligibility improvements obtained by use of the proposed MMSBA scheme employing diverse sub-band processing. Initial results have been very encouraging and will be reported elsewhere.

5. ACKNOWLEDGEMENTS

This work is supported by the U.K. Engineering and Physical Sciences Research Council (EPSRC) Project Reference Number GR/K48907.

6. REFERENCES

[1] R. Le Bouquin and G. Faucon, *Study of a voice activity detector and its influence on a noise reduction system,* Speech Communication, Vol.16, pp.245-254, 1995.

[2] E.Toner, *Speech Enhancement using Digital Signal Processing,* PhD thesis, University of Paisley, U.K. 1993.

[3] D.R.Campbell and E.Toner, *Speech enhancement with sub-band processing in an automobile environment,* 26th International Symposium on Automative Technology and Automation, Dedicated Conference on Mechatronics, Aachen, Germany 13th-17th September 1993.

[4] D.Darlington and D.R.Campbell, *Effect of modified filter distribution on sub-band adaptive speech enhancement scheme,* Proceedings EUSIPCO, Triest, Italy, 1996.

[5] H.S.Dabis, T.J.Moir and D.R.Campbell, *Speech enhancement by recursive estimation of differential transfer functions,* Proceedings of ICSP, Beijing, pp.345-348, 1990.

[6] H.S.Dabis and A.Wrench, *An evaluation of adaptive noise cancelling for speech recognition,* Proceedings EUROSPEECH, pp.1301-1304, 1991.

[7] A.Hussain, D.R.Campbell and T.J.Moir, *A Multi-microphone Sub-band Adaptive Speech Enhancement System employing diverse sub-band processing,* Proceedings ESCA-NATO Workshop, France, 17-18 April 1997.

[8] R.A.Goubran, R. Herbert and H.M.Hafez, *Acoustic noise suppression using regressive adaptive filtering,* 40th Vehicular Technology Conference, USA, pp.48-53, 1990.

[9] M.M.Goulding and J.S.Bird, *Speech enhancement for mobile telephony,* IEEE Trans. On Vehicular Technology, Vol.39 no.4, pp.316-326, 1990.

[10] E.Toner and D.R.Campbell, *Speech Enhancement using sub-band intermittent adaption,* International Journal of *Speech Communication,* Vol.12, pp.253-259, 1993.

[11] R.B.Wallace and R.A.Goubran, *Improved tracking adaptive noise canceller for non-stationary environments,* IEEE Trans. on Sig. Proc., Vol.40, no.3, pp.700-703, 1992.

[12] H.Yamada, H.Wang and F.Itakura, *Recovering of broad band reverberant speech signal by sub-band MINT method,* ICASSP, Toronto, Canada, pp.969-972, 1991.

[13] J.B.Allen, D.A.Berkley and J.Blauert, *Multi-microphone signal processing technique to remove room reverberation from speech signals,* J. Acoustic Soc. Amer., Vol.62, No.4, pp.912-915, 1977.

[14] E. R. Ferrara, B. Widrow, *Multi-channel Adaptive Filtering for signal enhancement,* IEEE Trans. on Acoustics, Speech and Signal Proc., Vol.29, no.3, pp.766-770, 1981.

[15] B.Widrow, S.D.Stearns, *Adaptive Signal Processing,* Prentice-Hall, 1985.

[16] Z.R. Zelinski, *Noise reduction based on microphone array with LMS adaptive post filtering,* Electronic Letters, Vol.26, No.24, pp.2036-2037, 1990.

[17] O.Ghitza, *Auditory models and human performance intasks related to speech coding and speech recognition,* IEEE Trans. Speech and AudioProc.,Vol.2, pp.115-132,1994

[18] D.D.Greenwood, *A cochlear frequency-position function for several species-29 years later,* J. Acoustic Soc. Amer, Vol.86, No.6, pp.2592-2605, 1990.

[19] J.B.Allen, *How do humans process and recognise speech?,* IEEE Trans. Speech and Audio Proc., Vol.2, No.4, pp.567-577, 1994.

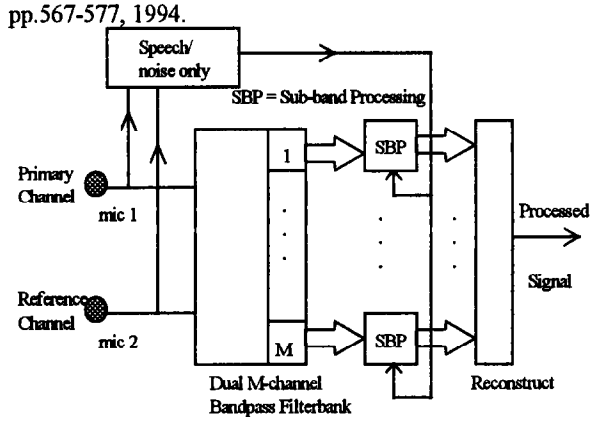

Figure 1: The proposed MMSBA scheme based on human cochlear model

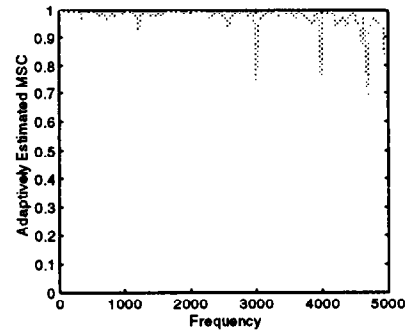

Figure 2: Adaptively estimated MSC between reverberant noises