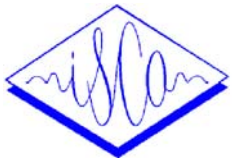

EUROSPEECH '97
5th European Conference on Speech Communication and Technology
Rhodes, Greece, September 22-25, 1997

A NEW METRIC FOR SELECTING SUB-BAND PROCESSING IN ADAPTIVE SPEECH ENHANCEMENT SYSTEMS

Amir Hussain, Douglas R. Campbell and Thomas J. Moir
Department of Electronic Engineering and Physics,
University of Paisley, High St., Paisley PA1 2BE, Scotland U.K.
Corresponding author's email: huss_ee0@paisley.ac.uk

10.21437/Eurospeech.1997-409

ABSTRACT

A multi-microphone adaptive speech enhancement system employing diverse sub-band processing is presented. A new robust metric is developed, which is capable of real-time implementation, in order to automatically select the best form of processing within each sub-band. It is based on an adaptively estimated inter-channel Magnitude Squared Coherence (MSC) relationship, which is used to detect the level of correlation between in-band signals from multiple sensors during noise-alone periods in intermittent speech. This paper reports recent results of comparative experiments with simulated anechoic data extended to include simulated reverberant data. The results demonstrate that the method 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 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 in fact, considered quite acceptable to human listeners [7]. 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 [6], however performance deteriorates in reverberant environments. Multi-band processing has been found to be important in combating reverberation effects [10] [11]. Adaption is necessary to compensate for changing noise fields [2] [10] due to for example, non-Gaussian sources, source/sensor motion, or time-varying acoustic paths. Multi-sensor methods are necessary to compensate for reverberation [13] and speech/noise spectral overlap [2].

Studies on noise in automobiles by various researchers [9] [10] [11] have found that the long-term noise correlation between two microphone locations was high for frequencies below 500Hz, and decreased gradually with virtually no correlation above 2kHz. Wallace and Goubran [12] 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.

Van Compernolle et al [5] and Dabis et al [6] 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] [8] [11]. 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.

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 [4]. 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.

2. THE PROPOSED MMSBA SYSTEM INCORPORATING THE NEW METRIC

Two or more relatively closely spaced microphones may be used in an adaptive noise cancellation scheme [4] [5] [6] [11] to identify a differential acoustic path transfer function during a noise only period in intermittent speech. The extension of this work, termed the Multi-Microphone sub-band Adaptive (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 either 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.

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 set the processing transfer function to one, that is, 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 it is often possible to arrange the sensor placement to avoid this acoustic path balancing.

3. Incoherent noise canceller: If the noise power is significant but not highly correlated between the two channels in a sub-band, then the incoherent noise cancellation approach of Ferrara and Widrow [15] or Zelinski [17] may be applied during the noisy speech period. Since in this case, the primary signal noise component $BH_1 N$ is uncorrelated with the reference signal noise component $BH_2 N$, the filtered reference is an estimate of the speech signal *S*.

In this paper, we examine the above three SBP options and implement the processing using the Least Mean Squares (LMS) algorithm [16] to perform the adaption. A metric for selecting the appropriate SBP option is now derived next.

2.1 The adaptively estimated Magnitude Squared Coherence (MSC) Metric

The coherence function is a complex function of frequency defined as [14]:

$$\rho(f) = \frac{E[Y(f)Z^*(f)]}{\sqrt{E\left[|Y(f)|^2 |Z(f)|^2\right]}}$$

where $Y(f)$ and $Z(f)$ are the Fourier transforms of the signals $y(n)$ and $z(n)$ picked up at the same time by two microphones. The magnitude of the complex coherence function varies between 0 and 1 and gives, for each frequency, the percentage of signal energy coming from correlated sources; with a value close to 1 implying strong correlation between the two signals and a value close to 0 indicating uncorrelated (or weakly correlated) signals.

In the context of dereverberation, the Magnitude Squared Coherence (MSC) has been previously used by Allen et al [14] to correct the magnitude of the reflected signal. Recently, Bouquin and Faucon [1] have applied the MSC on noisy speech signals 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.

Since the speech signal uttered by the speaker is submitted to modifications due to its propagation, the observations received by the two microphones, mic 1 and mic 2, as shown in Figure 1, may be written as (assuming that the speech and noise signals are independent):

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 f_k, the coherence function is given by [13]:

$$\rho(f_k, l) = \frac{S_{x_1 x_2}(f_k, l)}{\sqrt{S_{x_1 x_1}(f_k, l) S_{x_2 x_2}(f_k, l)}}$$

where $S_{x_1 x_2}(f_k, l)$ is the cross-spectral density, $S_{x_1 x_1}(f_k, l)$ and $S_{x_2 x_2}(f_k, l)$ are the auto-spectral densities; which can be estimated using a simple recursive calculation on a block by block basis [14]:

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

where β is the 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 f_k, $k = 0,...,L/2$, (where L corresponds to the length of the short term FFT and is set to L=256) as:

$$\mathrm{MSC}(f_k, l) = \frac{|S_{x_1 x_2}(f_k, l)|^2}{S_{x_1 x_1}(f_k, l) S_{x_2 x_2}(f_k, l)}$$

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

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

The above *adaptively averaged* MSC criterion can thus be used as an effective means for determining the level of correlation between the disturbing noises at various frequencies 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 on the basis of the inter-channel correlation.

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 and Faucon [1].

3. SIMULATION RESULTS

3.1 Simulated Anechoic Data:

Two simulated independent white noise sequences with an averaged Mean Squared Coherence relationship depicted in Figure 2 were used to corrupt a real anechoic speech signal sampled at 10kHz. The initial SNR was fixed at -2dB, and a noise alone period was manually labelled comprising the first 1024 samples. 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 arbitrarily chosen to be 256.
2. A two sensor MMSBA system with four sub-bands employing intermittent adaptive LMS update in each of the 4 sub-bands (MBLMS). The sub-band filter order was set to 256/4.
3. And the proposed MMSBA system also with four sub-bands, employing diverse sub-band processing (MBDLMS) with the intermittent adaptive LMS update employed in each of the first two bands (in order to more effectively cancel the correlated noises in those sub-bands cf. Figure 2) and the continuous Ferrara-Widrow LMS update was employed in each of the last two bands (in order to more effectively cancel the uncorrelated noises in those sub-bands cf. Figure 2).

System	FBLMS	MBLMS	MBDLMS
SNR improv.	1.8dB	5.7dB	10.8dB

Table 1: Performance Comparison of various adaptive noise cancellers for synthetic data.

As can be seen from Table 1, for this test case the use of a MBLMS system gives better performance in cancelling the simulated interference compared to the conventional FBLMS noise canceller. However, the use of the proposed MBDLMS system employing diverse sub-band processing, dependent on the inter-channel coherence information, can be seen to produce a much greater performance increment over the MBLMS system. Informal listening tests also showed the MBDLMS processed speech to be both enhanced in SNR and of better perceived quality than that obtained by the other methods.

3.2 Simulated Reverberant Data:

For this case, the impulse responses between the noise source and two microphones were calculated by an image program 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 approximated to 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, 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 the previous 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 across all the frequency bands. Thus for this case, the MBDLMS noise canceller is identical to the MBLMS canceller. The performance of the FBLMS and the 4-band MBDLMS intermittent noise cancellers is compared in Table 2.

System	FBLMS	MBDLMS
SNR improv.	1.9dB	10.1dB

Table 2: Performance Comparison of adaptive noise cancellers for "realistic" room data.

As can be seen from Table 2, the use of a MBDLMS system gives significantly better performance in cancelling the simulated room reverberated noise compared to the conventional FBLMS noise canceller. Informal listening tests again confirmed the performance improvements.

4. CONCLUSIONS

A multi-microphone sub-band adaptive (MMSBA) speech enhancement system employing diverse sub-band processing has been presented. An adaptively estimated inter-channel MSC measure has been proposed for selecting the best form of processing within each sub-band. Comparative results achieved in simulation experiments demonstrate that the method is capable of outperforming conventional noise cancellation schemes. Current experiments are extending the work reported here by using anechoic speech signals corrupted with reverberated noises recorded in real room and automobile environments.

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, Department of Electronic Engineering and Physics, University of Paisley, 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] 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 on *Robust Speech Recognition for Unknown Communication Channels*, pp.123-126, Pont-a-Mousson, France, 17-18 April 1997

[5] D.Van Compernolle, W. Ma and M.M. Van Diest, *Speech recognition in noisy environment with the aid of microphone arrays*, European Conf. On Speech Technology, Vol.2, pp.657-660, Paris, France, 1989.

[6] H.S.Dabis, T.J.Moir and D.R.Campbell, *Speech enhancement by recursive estimation of differential transfer functions*, Proceedings of International Conference on Signal Processing (ICSP), pp.345-348, Beijing, China, 1990.

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

[8] D.Darlington and D.R.Campbell, *Sub-band Adaptive Filtering Applied to Speech Enhancement*, ESCA Research Workshop, The Auditory Basis of Speech Perception, Keele University, 15-19 July, 1996.

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

[10] 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.

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

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

[13] 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.

[14] 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.

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

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

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

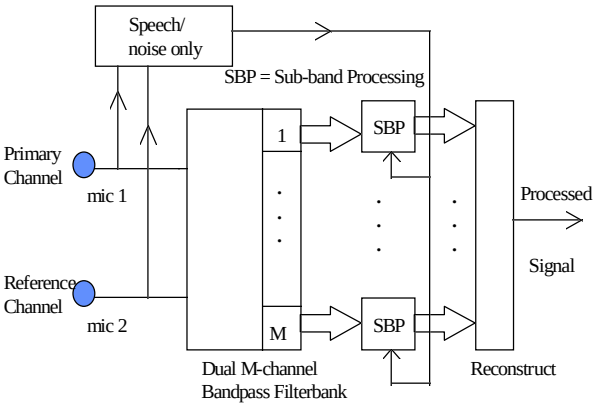

Figure 1: The MMSBA system

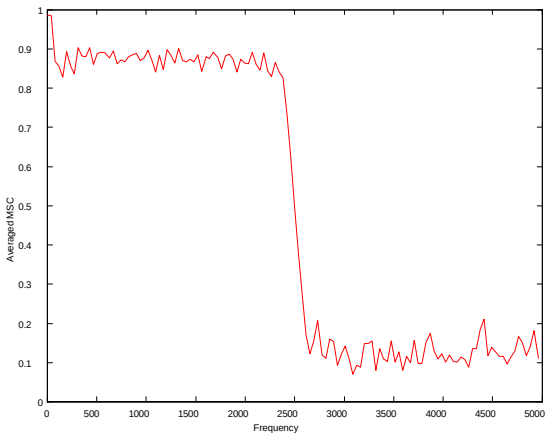

Figure 2: The adaptively estimated MSC between the disturbing synthetic noises