Speech Communication 12 (1993) 253–259
North-Holland

Speech enhancement using sub-band intermittent adaption

E. Toner and D.R. Campbell

Department of Electrical Engineering, University of Paisley, High Street, Paisley, Renfrewshire PA1 2BE, Scotland, UK

Received 16 February 1993

Abstract. A sub-band multisensor structure using intermittent adaption is proposed for speech enhancement. The convergence of the proposed method is compared with conventional LMS and frequency domain LMS and a dramatic increase in convergence rate is shown using both simulated and real data. Preliminary investigation of sub-band filter order is also reported.

Zusammenfassung. In diesem Artikel schlagen wir eine Mehrbandstruktur mit mehreren Gebern vor, die einen Mechanismus der intermittierenden Anpassung benutzt. Die Konvergenz dieser Methode wird mit der der Methode der geminderten Potenzen verglichen, die im zeitlichen und wiederholenden Bereich angewandt wird. Unsere neue Methode gewährleistet auch eine wesentliche Verbesserung der Konvergenz für simulierte und reelle Daten. Vorergebnisse über die Beeinflussung in der Reihenfolge der Filter in jedem Band sind ebenfalls dargestellt.

Résumé. Dans cet article nous proposons une structure multibande à plusieurs capteurs qui utilise un mécanisme d'adaptation intermittente. La convergence de cette méthode est comparée avec celle de la méthode des moindres carrés appliquée dans le domaine temporel et fréquentiel. Notre nouvelle méthode permet d'obtenir une amélioration très importante de la convergence pour des données simulées et réelles. Des résultats préliminaires concernant l'influence de l'ordre des filtres dans chaque bande sont aussi présentés.

Keywords. Speech enhancement; adaptive processing; multi-sensor; sub-band processing

1. Introduction

The enhancement of speech degraded by background noise by which we will mean an increase in signal-to-noise ratio (SNR), may be required to improve intelligibility for either human or machine recognition. Compared with humans, modern speech recognition equipment performance degrades markedly in the presence of background noise. In a recent experiment (Dabis and Wrench, 1991), the phoneme recognition rate fell from 92 to 71% correct under noise levels typical of a normal office environment (SNR $\approx$ 12 dB) in which human listeners would function without problems.

Some researchers have looked to the human hearing system as a source of engineering models to approach the enhancement problem, from Ghitza (1988) modelling the cochlea to recent work by Cheng and O'Shaughnessy (1991) utilising a model of the lateral inhibition effect. A recurring feature in this body of work is the accepted model of the cochlea as a spectrum analyser.

Single channel enhancement strategies generally suffer when the noise spectrum overlaps that of the speech. Humans can function well in such circumstances, as shown by the "cocktail party" effect, which can be partly attributed to multisensor usage since performance degrades with sensory path damage. The existence of the "binaural unmasking" effect (Evans, 1982) supports the use of multiple sensors for noise reduction as well as spatial localisation, appearing functionally equivalent to Widrow's classic noise cancelling (Widrow and Stearns, 1985). Further complicat-

ing the enhancement problem is the non-stationarity of many everyday noise sources and the effects of room acoustics. Humans may invoke short term adaption strategies perhaps related to the effect reported by Summerfield et al. (1984), to compensate for these.

The work reported here proposes a sub-band multi-sensor structure for speech enhancement which incorporates an intermittent adaptive process.

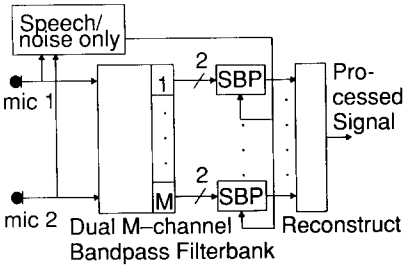

Fig. 2. Proposed 2-mic. processing scheme.

2. Proposed scheme

Two or more relatively closely spaced microphones may be used (Van Compernolle and Van Diest, 1989; Dabis et al., 1990; Campbell et al., 1992) to identify a differential acoustic path transfer function during a noise only period in intermittent speech. The locations of the two microphones and the noise source within a room will produce two acoustic transfer functions H_1 and H_2 (see Figure 1), a function of which can be identified using an adaptive algorithm. This function may then be used during the speech period (assuming short term constancy) to process the noisy speech signal. The extension of this work applies the method within a set of sub-bands provided by a filterbank as in Figure 2.

The speech/noise only detector is assumed available (e.g. (Tucker, 1992)) and although not a trivial problem will not be considered further here. In the following work noise only sections were manually identified.

The filter bank could be obtained by various orthogonal transforms or by a parallel filter bank approach. While the latter has considerable advantage in practical implementation a readily available Fast Hartley Transform (Bracewell, 1984) was used here. For these initial trials constant bandwidth band-pass filters were used.

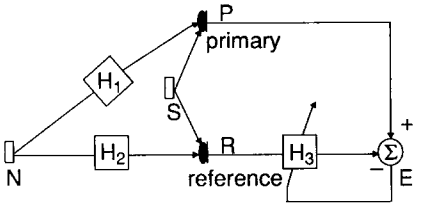

Fig. 1. Signal process configuration, after Dabis et al. (1990).

The sub-band processing (SBP) could be accomplished in a number of ways. For example,

(i) Examine the noise power and if below some threshold set the processing transfer function to unity.

(ii) If the noise power is significant and the noise is significantly correlated between the two channels, then perform adaptive noise cancelling.

(iii) If the noise power is significant but not highly correlated between the two channels, then use the adaptive cancelling approach of Zelinski (1990). This latter option has been included since we have found the noise to exhibit different levels of correlation between the two channels in different frequency bands.

We are presently examining the last two options and implementing the processing using the Least Mean Square (LMS) algorithm (Widrow and Stearns, 1985) to perform the adaption. This processing is based on the model of Fig. 1, where it is assumed for simplicity that the speaker is close enough to the microphones that room acoustic effects on the speech are insignificant and 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.

Referring to Fig. 1 N, S, P, R represent the z-transforms of the noise signal, speech signal, primary signal and reference signal, respectively. Thus at the primary

$$P = S + H_1 N, \tag{2.1}$$

and at the reference

$$R = S + H_2 N; \tag{2.2}$$

therefore

$$E = (1 - H_3)S + (H_1 - H_3 H_2)N. \tag{2.3}$$

The noise cancelling problem is to find H_3 such that the variance J_e of the error is minimised,

$$J_e = \frac{1}{2\pi j}\oint_{|z|=1} EE^* z^{-1}\, dz, \tag{2.4}$$

and during a noise only period $S = 0$, defining the noise spectral density Θ_{nn}, then

$$J_e = \frac{1}{2\pi j}\oint_{|z|=1} (H_1 - H_3 H_2) \times \Theta_{nn}(H_1 - H_3 H_2)^* z^{-1}\, dz, \tag{2.5}$$

which is minimised in the least squares sense when

$$H_3 = H_1 H_2^{-1}, \tag{2.6}$$

which is a transfer function that minimises the noise appearing in E.

Now using H_3 as a fixed processing filter when speech and noise are present *ideally* yields

$$E = (1 - H_3)S, \tag{2.7}$$

which is a noise reduced, filtered version of the speech signal.

3. Implementation of proposed scheme

Bandpass filtering was performed using a Fast Hartley Transform producing a set of signals in M contiguous frequency bands allowing independent sub-band processing to be applied.

The results presented here use Widrows LMS algorithm in all sub-bands. The convergence constant μ of the LMS algorithm was calculated for each individual band dependent on the variance of the signal within each band (Narayan and Peterson, 1981). Some researchers estimate sub-band filter order as either (i) the length of the conventional LMS filter divided by the number of bands (Somayazulu et al., 1989), (ii) the length of the conventional LMS filter divided by the down-sampling factor L (Hatty, 1990), (iii) the length of the echo signal within the corresponding band (Gilloire, 1987). We use the estimate given by (i), however, early indications show that it is more likely that different sub-bands may require different order filters.

Once the adaption process is stopped, the sub-band adaptive filters ($H_{f_1} \ldots H_{f_M}$) are fixed and the filtering process takes place during the speech plus noise period.

Preliminary results using the other possible sub-band processing are encouraging but require further investigation.

4. Results

4.1. Introduction

The performance of the proposed method defined as multi-band LMS (MBLMS) was compared with the established methods of frequency domain LMS (FDLMS) (Widrow and Stearns, 1985; Lee and Un, 1986) and conventional time domain LMS (CLMS) (Widrow and Stearns, 1985) by examining the convergence of their mean square errors (mse), see Figures 3–5, respectively. The effect on mse of varying sub-band filter length within each frequency band will also be shown. Results presented are firstly those obtained using simulated room data followed by those obtained using data recorded in a real environment. The noise source for the initial test was chosen to be

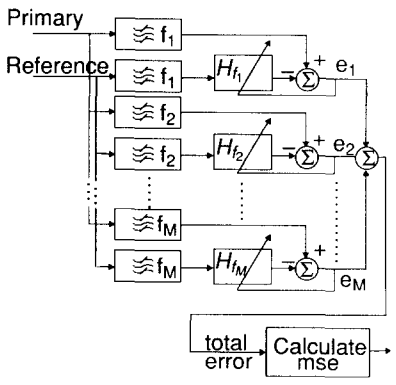

Fig. 3. Multiband LMS (MBLMS).

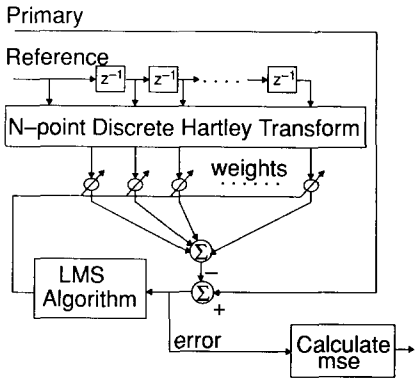

Fig. 4. Frequency domain LMS (FDLMS).

white noise as a simple method of injecting some noise power into each band.

4.2. Simulated room results

Test data was synthesized as shown in Figure 6.

The impulse responses between the noise source and the two microphones were calculated by a program which simulates room acoustics using room dimensions, reflection coefficients and source/receiver locations as parameters. Realistic responses would be of length > 1024 at a 10 kHz sampling rate but for testing purposes a length of 256 was selected. Two synthetic microphone signals were then generated by convolving a white noise sequence with each of the simulated impulse responses to yield the primary and reference signals. These were then used as the inputs to the three adaptive noise cancelling processes to be compared.

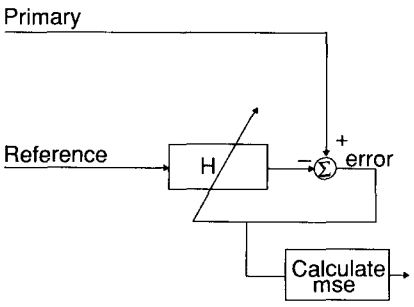

Fig. 5. Conventional LMS (CLMS).

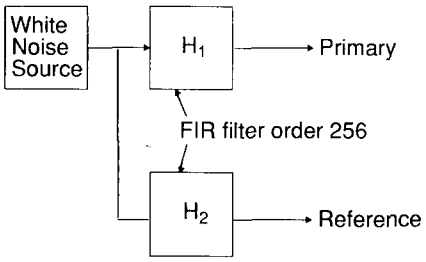

Fig. 6. Production of synthetic test data.

The established CLMS and FDLMS methods use a single error signal in the weight update vector. The MBLMS method minimises the error signal in each frequency band. For comparison purposes the mse was evaluated from a *single* total error output from each configuration. The summed error signal for the multiband approach was used to calculate the mse since the individual band-limited error signals are effectively orthogonal. This was verified by evaluating cross-product terms.

An adaptive filter length of 256 was set for CLMS to identify. In an attempt to equalise computational requirements the multiband method used a filter length of $256/M$ in each band (M is the number of sub-bands).

The mse convergence of all three methods is shown in Figure 7. When the number of fre-

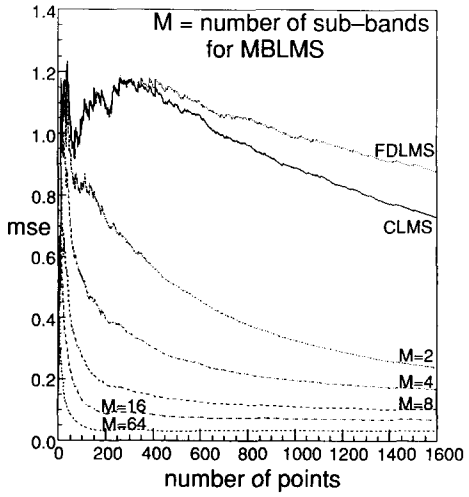

Fig. 7. MBLMS versus FDLMS versus CLMS for simulated room.

quency bands $M = 1$, the multiband method is obviously identical to CLMS. As M is increased (and adaptive filter length correspondingly decreased) the improvement in convergence speed is dramatic. FDLMS can have faster convergence than CLMS if the reference input is coloured noise. This allows for the pre-whitening effect of FDLMS (Narayan and Peterson, 1981) to increase convergence speed. However, for our test data the reference input is already a white noise signal, hence FDLMS and CLMS have similar convergence performance.

4.3. Real room results

The experimental set-up for recording real data was as shown in Figure 8.

Two microphones were placed centrally within a reverberant room spaced approximately 40 cm apart and 1 m distant from a loudspeaker driven by a white noise generator. The room dimensions were $6 \times 5 \times 4$ m containing the computer system as well as desks, cabinets, etc. The signal from the microphones were passed through pre-amplifiers and anti-aliasing filters with a cut-off frequency of 4 kHz and digitised at a sampling rate of 10 kHz. A filter length of 256 was assumed as an estimate of the room impulse response. The mse convergence performance of all three methods is shown in Figure 9. Also shown on the graph is the adverse effect on the performance of

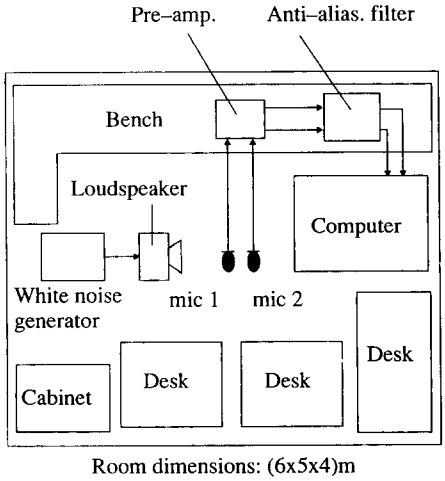

Fig. 8. Set-up for recording real data.

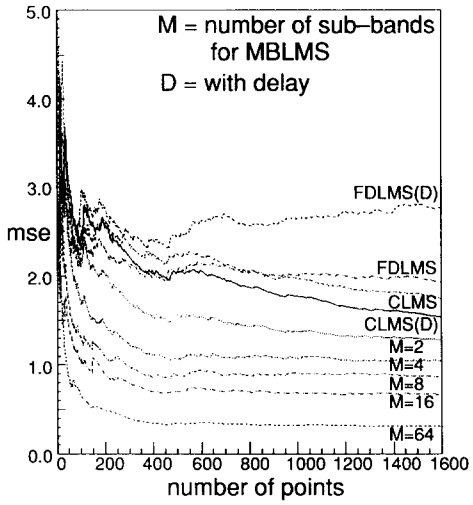

Fig. 9. MBLMS versus FDLMS versus CLMS for real room.

FDLMS of a delay between the signals received by the microphones. This delay has only a slight effect on CLMS and was evaluated using cross-correlation to be 12 samples at a sampling rate of 10 kHz. Compensating this delay moves the performance of FDLMS closer to that of CLMS in agreement with results presented by Reed and Feintuch (1981). All multiband results are with the delay present.

4.4. Effects of varying filter order within sub-bands

Varying the filter order used within each sub-band but keeping a fixed number of frequency bands reveals a trade-off between using low order filters and an increase in mse.

The real recordings of Section 4.3 also yielded results for 8- and 16-frequency bands, while an adaptive filter order in each of the bands was varied from 0 to 128. To compare performance for successive filter orders the value of the mse at the 500th data point was plotted against filter order as shown in Fig. 10. The actual mse value obtained by using $256/M$ as an estimate of filter length within each frequency band is indicated for both cases by the dashed line. The minimum mse value attained is indicated by the dotted line. For $M = 8$ the mse has reduced from 0.9 for order 32 to 0.75 for order 65. For $M = 16$ the mse has reduced from 0.73 for order 16 to 0.57 for

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.