

New Hybrid Sub-band Speech Enhancement Systems incorporating Neural Networks and post-Weiner Filtering

Amir Hussain
Department of Computing Science and Maths,
University of Stirling, FK9 4LA, Scotland, UK
E-mail: ahu@cs.stir.ac.uk

ABSTRACT

In this paper, two new hybrid sub-band systems are proposed which aim to combine neural network sub-band processing with post-Wiener filtering for adaptive speech-enhancement processing of noisy signals. The proposed hybrid architectures comprise an early auditory-processing modelling inspired Multi-Microphone Sub-band Adaptive (MMSBA) system incorporating neural-network based non-linear sub-band filters, integrated with post-Wiener filtering (WF) in order to further reduce the residual incoherent noise components resulting from the application of conventional non-linear MMSBA processing (without WF). A human cochlear model resulting in a non-linear distribution of the sub-band filters (as in humans) is also employed in the developed schemes. Preliminary comparative results achieved in simulation experiments using anechoic speech corrupted with real automobile noise show that the proposed structures are capable of significantly outperforming the conventional non-linear MMSBA and wide-band noise cancellation schemes.

1. INTRODUCTION

Speech enhancement is motivated by the need to improve the performance of voice communications systems in noisy conditions. The goal is either to improve the perceived quality of the speech, or to increase its intelligibility.

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 [1], however performance deteriorates significantly in reverberant environments. Typical results for hearing-aid applications achieve noise suppression of around 20dB in anechoic conditions dropping to 3dB to 6dB in reverberant surroundings [2].

Adaptive sub-band processing, which is inspired by cochlear mechanical filtering performed in the auditory periphery, has been found to overcome these limitations in general time-varying noise fields [2-5]. However, the type of processing for each sub-band must take effective account of the characteristics of the coherence between noise signals from multiple sensors. Several experiments have shown that noise coherence can vary

with frequency, in addition to the environment under test and the relative locations of microphones [2][4].

The above evidence implies that processing appropriate in one sub-band, may not be so in another, hence supporting the idea of involving the use of diverse processing in frequency bands, with the required sub-band processing being identified from features of the sub-band signals from the multiple sensors [4].

Dabis et al. [1] 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 inspired by certain features of early auditory processing modeling, such as cochlear mechanical filtering and binaural ‘unmasking’, has been described which extends this method by applying it within a set of linearly or non-linearly spaced sub-bands provided by a filter-bank [2][4][5]. In pilot studies, this non-optimised *linear* MMSBA scheme incorporating linear Finite Impulse Response (FIR) based sub-band filters, has shown the potential to yield up to 20dB signal-to-noise ratio (SNR) improvements over conventional wide-band (linear FIR filtering based) methods in real-reverberant room & automobile environments [2][4]. Recent pilot experiments using normal and hearing-impaired human-listeners and real-noisy reverberated speech have demonstrated statistically-significant improvements in intelligibility & perceived-quality [5]. Also note that the MMSBA scheme assumes noisy speech input to both (or all) system sensors, in contrast to the practically restrictive ‘classical’ full-band speech enhancement schemes, where speech signal occurs only at the primary input sensor [2]. This makes the MMSBA solution more attractive for practical realization and extends the range of applications in which it can be employed. However an effective method for detecting noise-only periods is assumed available within the MMSBA schemes.

Preliminary experiments [6][7] with a similar structure but incorporating relatively low-complexity artificial neural networks (ANN) as novel non-linear sub-band processing elements (the overall structure termed *non-linear* MSSBA processing scheme) show significantly improved relative SNR performance (verified by informal listening tests) over the conventional linear MSSBA and wide-band schemes in a real reverberant automobile-environment. The superior performance

of the non-linear MMSBA scheme is attributed to the incorporated ANN based sub-band filters which are better capable of taking account of the non-Gaussian nature of speech and non-linear distortions in electro-acoustic transmission systems [6]. The non-linear MMSBA framework attempts to conceptually combine cochlear mechanical filtering (performed in the auditory periphery), with a form of neural-network sub-band processing to approximate the neural-circuits in the auditory brainstem.

In this paper, the novel use of post-Wiener filtering (WF) within the non-linear MMSBA scheme is investigated, in order to more effectively deal with residual incoherent noise components that may result from application of the conventional non-linear MMSBA scheme (without WF). This preliminary work also extends that recently reported in [8] where a *linear* sub-band adaptive noise-cancellation scheme utilizing WF was developed for the *monaural* case. Performance of the two proposed *hybrid* non-linear MMSBA (incorporating post-WF) schemes is compared with the stand-alone non-linear MMSBA scheme (without WF) both quantitatively and qualitatively using informal subjective listening tests, for the case of a real anechoic speech signal corrupted with simulated noise, and initial results appear promising.

The paper is organized as follows: the proposed non-linear MMSBA schemes incorporating WF are described in section 2, including the choice of neural-network based Sub-band Processing (SBP), post-Weiner Filtering theory, details of the diverse SBP options available to the designer, and the adaptive correlation metric (CM) developed for selecting the appropriate SBP option. In section 3, preliminary simulation results are used to demonstrate the effectiveness of the proposed approach. Finally, some concluding remarks are presented in section 4.

2. New non-linear MMSBA Schemes incorporating WF

Two or more relatively closely spaced microphones may be used in an adaptive noise cancellation scheme [1] 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. The filter bank can be implemented using various orthogonal transforms or by a parallel filter bank approach. In this work, the sub-bands are distributed non-linearly according to a cochlear distribution, as in humans, following the Greenwood [9] model, in which the spacing of the sub-band filters is given by:

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

where x is the proportional distance from 0 to 1 along the cochlear membrane 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 recommended and chosen here.

The conventional *linear* MMSBA approach has been shown to considerably improve the mean squared error (MSE) convergence rate of an adaptive multi-band linear FIR filter compared to both the conventional wideband time-domain

and frequency domain FIR filters [2][4]. The use of a cochlear distribution of the sub-band filters as above, has also been shown to result in an equalized power distribution across the sub-bands (for the case of speech signals), resulting in further improved sub-band filter convergence compared to the case of linearly distributed sub-band filters [4][5].

The recently developed *non-linear* MMSBA scheme incorporating neural network based non-linear FIR (NLFIR) sub-band filtering is depicted in Figure 1, which is a further extension of the linear MMSBA approach, and has been shown to offer further performance benefits in preliminary studies [6][7].

Note that as depicted in Figure 1, we again assume 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 modelled as a point source modified by two different acoustic path transfer functions H_1 , H_2 , and that an effective voice activity detector (VAD) is available (as shown in Figure 1).

In the proposed *hybrid* non-linear MMSBA architecture shown in Figure 2, post-Weiner Filtering (WF) operation can be applied in two different ways: at the output of each sub-band processor (SBP), as shown in Figure 2a(i), or at the global output of the non-linear MMSBA scheme as shown in Figure 2a(ii). In the rest of this paper, the new non-linear MMSBA scheme employing WF in the sub-bands is termed *MMSBA-WF*, whereas the proposed non-linear MMSBA scheme employing wide-band (WB) WF is termed *MMSBA-WBWF* respectively.

In both the proposed hybrid architectures, the role of post-WF is to further mitigate the residual noise effects on the original signal to be recovered, following application of conventional non-linear MMSBA noise-cancellation processing.

The next sub-sections discuss: the non-linear Artificial Neural Network (ANN) based NLFIR filters used in SBP together with the new post-WF extensions, the choice of diverse SBP options and finally the Correlation Metric (CM) used for selecting the appropriate SBP option.

2.1 Artificial Neural Network (ANN) based SBP

A class of general adaptive non-linear FIR (NLFIR) type filters based on single hidden-layered linear-in-the-parameters ANNs is described in [6] for processing the band-limited signals in a multi-band speech enhancement system.

The general structure of the NLFIR type filter is based on single-hidden layered, linear-in-the-parameters feedforward ANNs, as shown in Figure 3. It employs an input expander which transforms the n inputs $[x_1, \dots, x_n]$ (representing lagged values of the sub-band input signal x passed through a tapped delay line of order $(n-1)$) into a non-linear intermediate (hidden) space of increased dimension N . The expanded input terms (termed the basis functions) are then weighted and linearly combined to form the adaptive filter output y . The overall mapping of the adaptive NLFIR is thus $\mathbf{R}^n \rightarrow \mathbf{R}^N \rightarrow \mathbf{R}$.

The advantage of this particular non-linear filter structure is

that linear adaptive filter theory can be readily applied for on-line adaptation [6]. The non-linear expansion model is completely general and can employ any of the non-linear basis functions commonly employed in e.g. the Radial Basis Function (RBF) neural networks, such as the thin-plate spline basis functions, multi-quadratic activation functions, the inverse multi-quadratic functions, or indeed the widely used Gaussian basis functions. Alternatively, the sigmoidal basis functions employed in Multi-Layered Perceptron (MLP) networks, or the Volterra (polynomial) expansion employed in the hidden layer of the conventional Volterra Neural Network (VNN) can also be employed. Another possibility described in [6] includes the *hybrid* functional-expansion employed in a recently developed Functionally-Expanded Neural Network (FENN), which is a variant of the conventional Functional-Link Neural-Network. The FENN expansion model comprises a *combination* of sigmoidal-shaped, Gaussian-shaped and polynomial-subset activation (basis) functions, and an additional benefit of this approach (like the VNN's polynomial-expansion) is that the use of the original network inputs within the expansion model, also enables efficient modeling of linear dynamical transfer-functions.

As discussed in [6], the choice of an appropriate non-linear expansion-model is, in general, problem dependent. For example, it has been shown that some problems such as functional approximation can be solved more efficiently with the sigmoidal-type basis functions employed in the MLP; while others such as classification problems are more amenable to localized (e.g. Gaussian-type) basis functions employed in the RBF. However, all the above expansion-models are known to be *universal approximators* as they can approximate any non-linear function to an arbitrary degree of accuracy.

Further details on the choice of an appropriate expansion model can be found in [6][7] where it has been argued that the relative performance-complexity trade-off for the non-linear models needs to be determined for each specific problem. However in practice, the simple polynomial expansion-model employed in the VNN is attractive since it requires relatively low-complexity hardware for implementation. In this paper, we shall restrict our choice of the NLFIR filter's expansion-model to be polynomial, *viz*:

$$\mathbf{f}(\mathbf{x}) = [1, x_{i1}, x_{i1} x_{i2}, \dots, x_{i1} x_{i2} \dots x_{ik}]$$

where $i\gamma = 1, \dots, n$ for $\gamma=1, \dots, k$ with k representing a k -th order polynomial expansion of the n (sub-band) filter inputs; and $\mathbf{f}(\cdot)=[f_1 \dots f_N]$ are the N non-linear basis functions.

Once the full expansion-model $\mathbf{f}(\mathbf{x})$ at the single hidden-layer of the ANN based NLFIR filter has been specified, conventional stochastic-gradient or least-squares based adaptation algorithms (such as the LMS, RLS or their robust versions) can then be used to provide an efficient means for real-time adaptation of the filter weights \mathbf{W} , as shown in Figures 2.b and 3 This will give these non-linear FIR filters a significant advantage over multi-layered (MLP type) neural-network based filters in recursive applications. Further details on the choice of adaptation algorithms for the ANN based NLFIR sub-band filters can be found in [6][7].

For the derivation of the post-WF theory in the next section, we now define $\tilde{X}_j, \tilde{S}_j, \tilde{N}_j$ as the global output, the reconstructed signal and the residual noise component at the j -th SBP output (or, equivalently, the adaptive NLFIR noise-canceller output of j band) respectively, as shown in Figure 2a. The following relationship can be assumed to hold due to noise and the desired signal at each band being uncorrelated:

$$\tilde{X}_j = \tilde{S}_j + \tilde{N}_j \quad (1)$$

In the original non-linear MMSBA (without WF), all \tilde{x}_j sub-band NLFIR noise canceller outputs are summed (at the reconstruction section) to yield the global MMSBA output \tilde{y} , which can be expressed in the frequency domain as:

$$\tilde{Y} = \sum_j \tilde{S}_j + \sum_j \tilde{N}_j = \tilde{S} + \tilde{N} \quad (2)$$

2.2 Post-Wiener Filtering (WF)

The coefficients of a Wiener filter (WF) [8] are calculated to minimise the average squared distance between the filter output and a desired signal, assuming stationarity of the involved signals. This can be easily achieved in the frequency domain yielding:

$$W(f) = (P_{DY}(f)/P_{YY}(f)) \quad (3)$$

where $D(f)$ is the desired signal, $\hat{S}(f) = W(f)Y(f)$ is the Wiener filter output, $Y(f)$ the Wiener filter input and $P_{YY}(f), P_{DY}(f)$ are the power spectrum of $Y(f)$ and the cross power spectrum of $Y(f), D(f)$ respectively. If we apply such a solution to the case where the global signal is given by addition of noise and signal (to be recovered), and moving from the assumption that noise, signal are uncorrelated (as \tilde{S}_j, \tilde{N}_j are) we can derive the following from [11]:

$$W_j(f) = \left(P_{\tilde{S}_j \tilde{S}_j}(f) / P_{\tilde{S}_j \tilde{S}_j}(f) + P_{\tilde{N}_j \tilde{N}_j}(f) \right) \quad (4)$$

where $P_{\tilde{S}_j \tilde{S}_j}(f), P_{\tilde{N}_j \tilde{N}_j}(f)$ are the signal and noise power

spectra. Note that, in this task, the desired signal is \tilde{S}_j .

It must be observed that such a formulation can be easily extended to the case when involved signals are not stationary, by simply periodically recalculating the filter coefficients for every block l of N_s signal samples. In this way the filter adapts itself to the average characteristics of the signals within the blocks and becomes block-adaptive.

Moreover, the presence of a VAD (in the MMSBA) is a prerequisite to making the Wiener filtering operation effective: in noise alone period, a precise estimation of noise power spectrum can be performed and then used in (4), assuming that its properties are still the same when the signal power spectrum is calculated during the noisy speech period. The former approximation is carried out iteratively by using the power spectrum of Wiener filter global output $\hat{S}_j(f)$.

Note that the above WF derivations are readily applicable to the

hybrid MMSBA-WF architecture as follows. Similar to (1) and (2), the following holds at the j -th band Wiener filter output:

$$\begin{aligned}\hat{X}_j &= \hat{S}_j + \hat{N}_j \\ \hat{Y} &= \sum_j \hat{S}_j + \sum_j \hat{N}_j = \hat{S} + \hat{N}\end{aligned}\quad (5)$$

where \hat{y} is the new global output yielded from the reconstruction section.

Finally, the same considerations can be made when the hybrid MMSBA-WBWF structure is dealt with, simply adapting the above equations to the new situation where WF occurs after the reconstruction section. Specifically, taking (2) into account, implies:

$$\hat{Y}_f = W_f \tilde{Y} = \hat{S}_f + \hat{N}_f \quad (6)$$

where f indicates full-band processing, since WF operation is applied directly to non-linear MMSBA output \tilde{y} to form the new Wiener filtered output \hat{y}_f .

Next, the choice of various non-linear SBP options available to the designer, are discussed.

2.3 Diverse SBP options

A significant advantage of using SBP for non-linear MMSBA speech enhancement is that it allows independent processing in each sub-band in an attempt to cancel the dominant noise components, coherent or non-coherent, present in each sub-band. The SBP can be accomplished in a number of ways (as depicted in Figure 2b), 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 the adaptive NLFIR 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 (assuming short term constancy) 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:

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

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 NLFIR adaptive filter. The Mean Squared Error function is:

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

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_m , then

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

which is minimised in the least squares sense when

$$H_3 = (B H_1)(B H_2)^{-1}$$

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

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

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

3. *Non-coherent noise canceller*: If the noise power is significant but not highly correlated between the two channels in a sub-band, then the non-coherent noise cancellation approach of Ferrara and Widrow (FW) [10] may be applied here during the noisy speech period. Since in this case, the primary signal noise component $B H_1 N$ is uncorrelated with the reference signal noise component $B H_2 N$, the filtered reference (output of NLFIR) is now an estimate of the sub-band speech signal S .

In this paper, we employ the above three SBP options and implement the adaptive sub-band processing using neural network based NLFIR type filters together with post-WF as described earlier. In the next section, we describe a metric for selecting the appropriate type of SBP option.

2.4 Correlation Metric (CM) for Selecting SBP

The Magnitude Squared Coherence (MSC) has been used by Bouquin and Faucon [12] who have applied it for the reduction of noise in speech signals and have also employed it as a Voice Activity Detector (VAD) for the case of spatially uncorrelated noises. In this work, we employ the MSC within a Correlation Metric (CM), as a part of a system for selecting an appropriate SBP option in the non-linear MMSBA speech enhancement system. Assuming that the speech and noise signals are independent, the observations received by the two microphones, as shown in Figure 2, may be written as:

Assuming that the speech and noise signals are independent, the observations received by the two microphones are:

$$x_p = s_p + n_p \quad \text{primary}; \quad x_r = s_r + n_r \quad \text{reference}$$

where $s_{p,r}$, $n_{p,r}$ represent the clean speech signal and the additive noise, respectively. For each block l and frequency bin fk , the coherence function is given by:

$$\rho(fk, l) = \frac{P_{X_p X_r}(fk, l)}{\sqrt{P_{X_p X_p}(fk, l) P_{X_r X_r}(fk, l)}}$$

where $P_{X_p X_r}(f_k, l)$ is the cross-power spectral density, $P_{X_p X_p}(f_k, l)$ and $P_{X_r X_r}(f_k, l)$ are the auto-power spectral densities; which can be estimated by:

$$P_{X_p X_r}(f_k, l) = \beta P_{X_p X_r}(f_k, l-1) + (1-\beta) X_p(f_k, l) X_r^*(f_k, l)$$

where β is a forgetting factor. During the noise alone period, for each overlapped and Hanning windowed block l we compute the Magnitude Squared Coherence (MSC) averaged over all the overlapped blocks (at each frequency bin) as:

$$\overline{\text{MSC}}(f_k) = \frac{1}{l} \sum_{i=1}^l [\rho(f_k, i)]^2$$

Finally the correlation metric (CM) is estimated for each linear or cochlear spaced sub-band s (over the appropriate frequency range f_p to f_q Hz) as:

$$CM(s) = \sum_{k=p}^q \overline{\text{MSC}}(f_k)$$

The above CM can thus be used as a means for determining the level of correlation between the disturbing noise sources within each sub-band during the "noise-alone" period in intermittent speech. On the basis of this CM, the subsequent form of NLFIR and post-Weiner filtering based processing in each respective frequency band can be selected as either the intermittent-coherent noise canceller or the non-coherent FW type noise canceller, provided the absolute and relative sub-band noise powers are above an experimentally determined threshold.

3. SIMULATION RESULTS

In this section the two new hybrid non-linear MMSBA-based WF approaches are compared to the original non-linear MMSBA approach (without WF) in order to investigate their relative effectiveness. For experimental purposes, a real anechoic speech signal $s(k)$ is used as the desired signal, whilst the noise signals $n_1(k), n_2(k)$ are chosen to be real stereo car noise sequences recorded in a Ferrari Mondial T (1991 Model), using an Audio Technica AT9450 stereo microphone mounted on a SONY DCR-PC3-NTSC video camera and a sampling frequency of 44.1 kHz. The noise sequences were manually added to the anechoic speech sentence to manufacture different SNR cases.

The value of the initial SNR, namely SNR_i , is used as a reference for the three SNR improvements calculated at the output of each of the speech enhancement structures under study, namely: the original non-linear MMSBA without WF (Figure 1), the new hybrid MMSBA-WF (Figure 2a(i)) and the hybrid MMSBA-WBWF (Figure 2a(ii)). Taking into account the non-correlation between noise and signal on the same channel, we can define the SNR at the output level as:

$$SNR_o(f) = \begin{cases} \left[\frac{P_{\hat{Y}\hat{Y}}(f) - P_{\hat{N}\hat{N}}(f)}{P_{\hat{N}\hat{N}}(f)} \right] \\ \left[\frac{P_{\hat{Y}\hat{Y}}(f) - P_{\hat{N}\hat{N}}(f)}{P_{\hat{N}\hat{N}}(f)} \right] \\ \left[\frac{P_{\hat{Y}_f \hat{Y}_f}(f) - P_{\hat{N}_f \hat{N}_f}(f)}{P_{\hat{N}_f \hat{N}_f}(f)} \right] \end{cases} \quad (7)$$

where all involved power spectra are related to signals described by (2), (5), and (6). Moreover it has to be said that $P_{\hat{N}\hat{N}}(f)$ is calculated over a sub-range of the noise alone period where noise cancellers are assumed to have converged, since this is the noise power spectrum expected to occur when the desired signal is present. On this basis, $P_{\hat{N}\hat{N}}(f)$ and $P_{\hat{N}_f \hat{N}_f}(f)$ are obtained from Wiener filtered versions for the two new hybrid schemes addressed (MMSBA-WF and MMSBA-WBWF).

In this work, the sub-bands are achieved by modifying the spectra of the FFT of the input signals, and the number of filters is therefore limited by the size of the FFT. The processing in each sub-band is performed using the ANN-based adaptive non-linear FIR-type filters.

Choices for various experimental parameter values were selected on a trial and error basis as: speech-signal number of samples corresponding to a 2s long speech sentence; noise signal number of samples (in the manually defined noise alone period) corresponding to 0.2s of car noise recording; number of iterations of WF operation: 5; number of cochlear-spaced sub-bands: 4; number of taps (order) of VNN-based NLFIR adaptive sub-band filters: 32. A truncated 2nd order polynomial expansion of the sub-band NLFIR filter inputs was employed comprising the actual sub-band filter inputs and their square terms which resulted in a total of 64 terms (basis functions) in each sub-band NLFIR filter.

The following SBP options were compared for each of the three non-linear MMSBA schemes.

Case (A) Intermittent only SBP: In the first experimental case study, the intermittent coherent noise-canceller approach is employed as the only SBP option in each band. Table 1 summarizes the results obtained using the three non-linear MMSBA approaches: from which it can be seen that the hybrid MMSBA-WF and the hybrid MMSBA-WBWF both deliver an improved SNR performance over the original non-linear MMSBA approach (without WF).

Case (B): Diverse (intermittent/FW) SBP: In this case the value of the adaptive CM is used to employ both intermittent and non-coherent (FW) SBP options, with the former option used in the first sub-band (with a high CM) and the latter in the other three bands (with a low CM). This is justified by the coherence characteristics of available stereo noise signal. It can be seen from Table 2 that the choice of sub-band WF (within the hybrid MMSBA-WF scheme) gives the best results in this case, due to its operation in the sub-bands, resulting in more effective noise cancellation in the frequency domain, compared to the hybrid MMSBA-WBWF scheme (employing wide-band WF processing) as well as the conventional non-linear MMSBA (without WF).

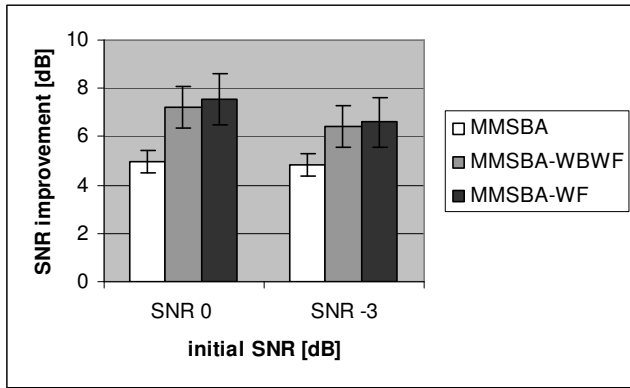


Table 1. Case (A): Comparison of various Non-linear MMSBA approaches (all adapted using intermittent SBP only). Relative average SNR improvements for all architectures involved (over 10 runs). Standard deviation values are directly depicted on bars.

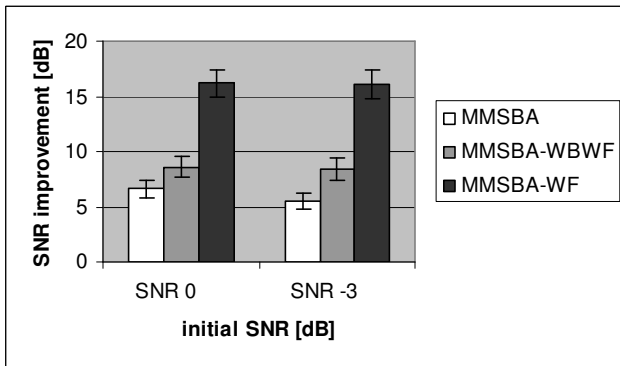


Table 2. Case (B): Comparison of various Non-linear MMSBA approaches all employing diverse (intermittent and non-coherent FW based) SBP. Relative average SNR improvements for all architectures involved (over 10 runs). Standard deviation values are directly depicted on the bars.

Note that application of the classical linear wide-band noise cancellation approach, namely the MMSBA with number of bands set to one and a wideband linear FIR filter order of 256 (of comparable complexity to the non-linear MMSBA) was actually found to degrade the speech quality resulting in a negative SNR improvement value, which is hence not shown in the Table 1. This finding of the inability of classical wideband processing to enhance the speech in real automobile environments is consistent with the results reported in [2][3].

Finally, informal listening tests using random presentation of the processed and unprocessed signals to three young male adults of normal hearing, also confirmed the MSSBA-WF processed speech to be both enhanced in SNR and of significantly better perceived quality than that obtained by all the other conventional wide-band and sub-band (non-linear MMSBA) methods.

4 Concluding Remarks

Two new hybrid multi-microphone sub-band adaptive (MMSBA) speech enhancement systems incorporating neural network based sub-band processing with post-Wiener filtering

and a human cochlear model function have been presented. Preliminary comparative results achieved in simulation experiments demonstrate that the proposed hybrid non-linear MMSBA processing schemes are capable of improving the output SNR of speech signals with no additional distortion apparent, compared to the conventional neural network based non-linear MMSBA scheme (without WF). The new hybrid MMSBA-WF architecture employing sub-band based post-WF seems to be the most promising whose superior performance can be attributed to the ability of the WF to further reduce the residual in-coherent sub-band noise components resulting from application of the conventional non-linear MMSBA scheme. A detailed theoretical analysis is now proposed to define the attainable performance. What is also needed is further extensive testing (using formal subjective listening tests) with a variety of real data (i.e., acquired through recordings in real environment), in order to further assess and quantify the relative advantages of the new speech enhancement schemes. Further work will also investigate the possibility of including cross-processes such as human lateral inhibition effects in the multi-band systems.

REFERENCES:

- [1] H.S. Dabis, T.J. Moir, and D.R. Campbell, "Speech enhancement by recursive estimation of differential transfer functions," *Proceedings of ICSP*, pp. 345-348, Beijing, 1990.
- [2] E. Toner, *Speech Enhancement using Digital Signal Processing*, PhD thesis, University of Paisley, UK, 1993.
- [3] Le Bouquin R, Azirani A & Faucon G: "Enhancement of speech degraded by coherent and incoherent noise using a cross-spectral estimator", *IEEE Trans. Speech & Audio Proc.* (SAP), Vol.5, no.5, 484-487.
- [4] A. Hussain, "A Multi-microphone Sub-band Adaptive Speech Enhancement System employing diverse sub-band processing," *International Journal of Robotics & Automation*, vol. 15, no. 2, pp. 78-84, 2000.
- [5] A.Hussain & D.R.Campbell, Intelligibility improvements using binaural diverse sub-band processing applied to speech corrupted with automobile noise, *IEE Proceedings: Vision, Image & Signal Processing*, Vol. 148, no.2, pp.127-132, 2001.
- [6] A.Hussain, Multi-sensor Neural Network processing of Noisy Speech, *International Journal of Neural Systems*, Vol.9, No.5, World Scientific Publ., UK, pp.467-472, 1999.
- [7] A.Hussain, Non-linear Speech Processing using Neural Networks based Adaptive Filtering, *Proc. 4th IEEE INMIC*, Islamabad, 10-11 Sep 2000
- [8] H. R. Abutalebi, H. Sheikhzadeh, R. L. Brennan, & G. H. Freeman, "A hybrid sub-band system for speech enhancement in diffused noise fields," *IEEE Sig. Process. Letters*, 2003.
- [9] 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.
- [10] E.R. Ferrara, and 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
- [11] S.V. Vaseghi, *Advanced signal processing and digital noise reduction (2nd ed.)*, John Wiley & Sons, 2000.
- [12] R. Le Bouquin & G. Faucon, "Study of a voice activity detector and its influence on a noise reduction system", *Speech Communication*, vol. 16, pp. 245-254, 1995.

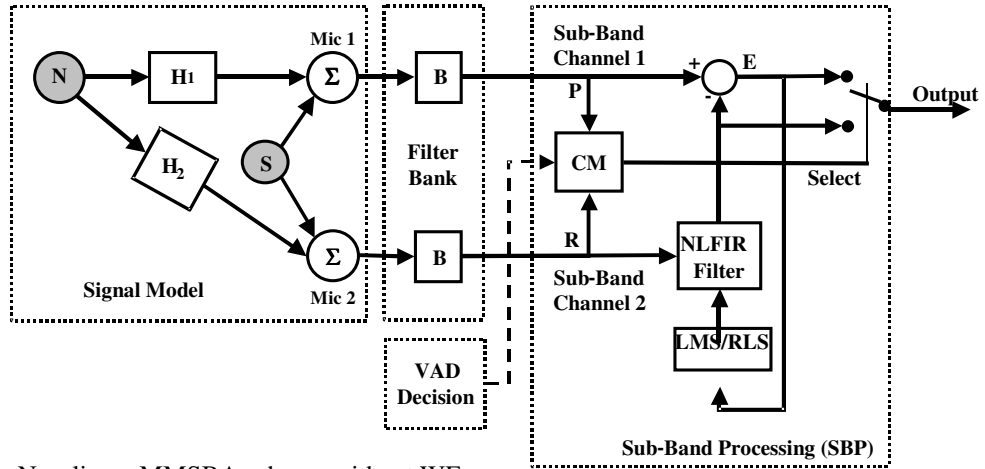


Figure 1: Non-linear MMSBA scheme without WF

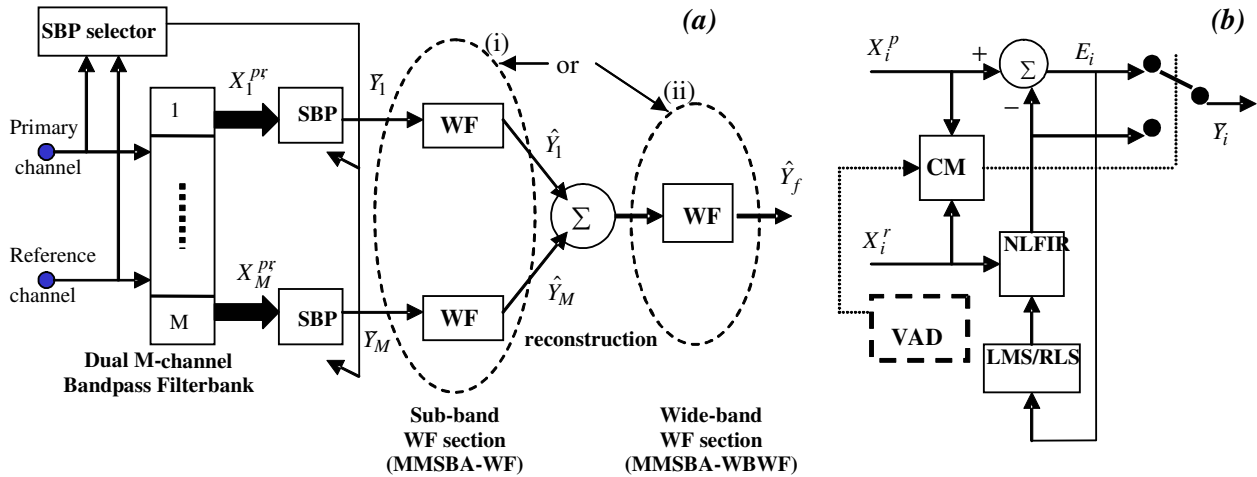


Figure 2 (a): New hybrid non-linear MMSBA systems incorporating post-Weiner Filtering (WF) in the form of: (i) MMSBA-WF, or, (ii) MMSBA-WBWF configuration

Figure 2 (b): Sub-band Processing (SBP) configurations

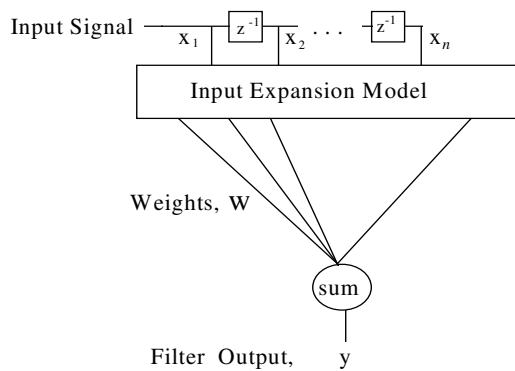


Figure 3: General structure of the proposed neural-network based adaptive non-linear FIR (NLFIR) type Filter used in SBP (Figure 2.b above)