Novel Sub-band Adaptive Systems Incorporating Wiener Filtering for Binaural Speech Enhancement

Amir Hussain[1], Stefano Squartini[2], and Francesco Piazza[2]

[1] Department of Computing Science & Mathematics, University of Stirling, Stirling FK9 4LA, Scotland, UK
ahu@cs.stir.ac.uk
http://www.cs.stir.ac.uk/~ahu/
[2] Dipartimento di Elettronica, Intelligenza Artificiale e Telecomunicazioni, Università Politecnica delle Marche, Via Brecce Bianche 31, 60131 Ancona, Italy
http://www.deit.univpm.it/

Abstract. In this paper, new Wiener filtering based binaural sub-band schemes are proposed for adaptive speech-enhancement. The proposed architectures combine a Multi-Microphone Sub-band Adaptive (MMSBA) system with Wiener filtering in order to further reduce the in-coherent noise components resulting from application of conventional MMSBA noise cancellers. A human cochlear model resulting in a non-linear distribution of the sub-band filters 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 MMSBA scheme without Wiener filtering.

1 Introduction

The goal of speech enhancement systems is either to improve the perceived quality of the speech, or to increase its intelligibility. Classical 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 in reverberant environments. Adaptive sub-band processing has been found to overcome these limitations [2] in general time-varying noise fields. 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. 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. 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 has been described which extends this method by applying it within a set of linearly spaced sub-bands provided by a filter-bank [2]-[4]. Nevertheless, it must be noted that the MMSBA scheme assumes

M. Faundez-Zanuy et al. (Eds.): NOLISP 2005, LNAI 3817, pp. 318 – 327, 2005.

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 [1]. This makes the MMSBA solution more practically realizable. However, a proper method for detecting noise-only periods is assumed available within the MMSBA scheme. In this paper, the novel use of Wiener filtering (WF) within a binaural MMSBA scheme is investigated, in order to more effectively deal with residual incoherent noise components that may result from the application of conventional MMSBA schemes. This work originally extends that recently reported in [5] where a sub-band adaptive noise cancellation scheme utilizing WF was developed for the monaural case. Performance of the proposed binaural WF based approach is compared with the conventional MMSBA scheme (without WF) quantitatively and qualitatively using informal subjective listening tests, for the case of a real speech signal corrupted with simulated noise.

2 MMSBA Schemes Employing WF

Two or more relatively closely spaced microphones may be used in an adaptive noise cancellation scheme [1], [3] 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 1a). 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 [6] model, in which the spacing of the sub-band filters is given by:

$$F(x) = A(10^{ax} - k). \tag{1}$$

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 MMSBA approach considerably improves the mean squared error (MSE) convergence rate of an adaptive multi-band LMS filter compared to both the conventional wideband time-domain and frequency domain LMS filters, as shown in [3][4]. 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 modelled as a point source modified by two different acoustic path transfer functions, and that an effective voice activity detector (VAD) is available.

In the proposed MMSBA architecture, Wiener filtering (WF) operation has been applied in two different ways: at the output of each sub-band adaptive noise canceller as shown in Fig.1a, and at the global output of the original MMSBA scheme as shown in Fig. 1b. In the rest of this paper, the new MMSBA scheme employing WF in the sub-bands is termed MMSBA-WF, whereas the one employing wide-band (WB) WF is termed MMSBA-WBWF. In both the proposed architectures, the role of WF is to further mitigate the residual noise effects on the original signal to be recovered, following application of MMSBA noise-cancellation processing.

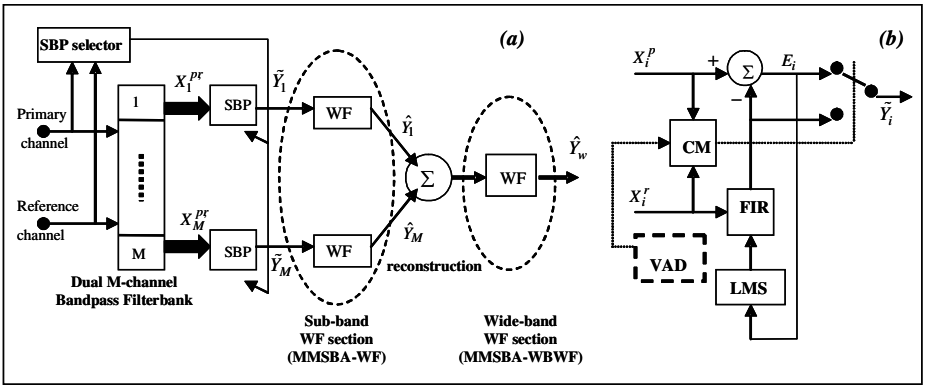


Fig. 1. (a) WF based MMSBA systems. (b) Subband Processing unit(SBP).

2.1 Diverse SBP Options

A significant advantage of using sub-band processing (SBP) for speech enhancement within the MMSBA scheme is that it allows for diverse processing in each sub-band in order to simultaneously effectively cancel both the coherent and incoherent noise components present in real reverberant environments. The SBP can be accomplished in a number of ways (Fig.1b), as follows:

- *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.
- *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 (assuming short term constancy) to process the noisy speech signal.
- *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-Widrow [7] be applied during the noisy speech period.

In this paper, we employ the above three SBP options and implement the processing using the Least Mean Squares (LMS) algorithm to perform the adaptation. For the derivation of the WF theory in the next section, we 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 noise canceller output of j band) respectively. The following relationship can be assumed to hold due to un-correlation between the noise and the desired signal at each band:

$$\tilde{X}_j = \tilde{S}_j + \tilde{N}_j. \tag{2}$$

In the original MMSBA, all $\tilde{x}_j$ sub-band noise canceller outputs are summed (at the reconstruction section) to yield the global MMSBA output $\tilde{y}$ (in the following capitalized letters will denote the corresponding variables in the frequency domain):

$$\tilde{Y} = \sum_j \tilde{S}_j + \sum_j \tilde{N}_j = \tilde{S} + \tilde{N}. \tag{3}$$

2.2 Wiener Filtering (WF)

The coefficient of a Wiener filter [8] are calculated to minimize 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:

$$W(f) = \left(P_{DY}(f) / P_{YY}(f) \right). \tag{4}$$

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 and signal are uncorrelated (as $\tilde{S}_j, \tilde{N}_j$ are) we can derive the following from (4):

$$W_j(f) = \left(P_{\tilde{S}_j \tilde{S}_j}(f) / P_{\tilde{S}_j \tilde{S}_j}(f) + P_{\tilde{N}_j \tilde{N}_j}(f) \right). \tag{5}$$

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 VAD is a pre-requisite 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 (5), 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 derivations are readily applicable to MMSBA-WF architecture as follows. Similar to (2) and (3), the following holds at *j*-th band Wiener filter output:

$$\hat{X}_j = \hat{S}_j + \hat{N}_j, \qquad \hat{Y} = \sum_j \hat{S}_j + \sum_j \hat{N}_j = \hat{S} + \hat{N} \tag{6}$$

where $\hat{y}$ is the new global output yielded from the reconstruction section. However, same considerations can be made when MMSBA-WBWF is dealt with, simply adapting the equations to the new situation where WF occurs after the reconstruction section. Specifically, taking (3) into account, implies:

$$\hat{Y}_w = W_w \tilde{Y} = \hat{S}_w + \hat{N}_w. \tag{7}$$

where *w* stands for wide-band processing, since WF operation is applied directly to MMSBA output $\tilde{y}$ to form the new Wiener filtered output $\hat{y}_w$.

2.3 Recursive Magnitude Squared Coherence (MSC) Metric for Selecting SBP

The Magnitude Squared Coherence (MSC) has been applied by Bouquin and Faucon [9] to noisy speech signals for noise reduction and also successfully employed as a VAD for the case of spatially uncorrelated noises. In this work, following [4] we use a modified MSC as a part of a system for selecting an appropriate SBP option within the MMSBA system. Assuming that the speech and noise signals are independent, the observations received by the two microphones are:

$$x_p = s_p + n_p \quad primary; \qquad x_r = s_r + n_r \quad reference. \tag{8}$$

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

$$\rho(f_k, l) = P_{X_p X_r}(f_k, l) \Big/ \sqrt{P_{X_p X_p}(f_k, l) P_{X_r X_r}(f_k, l)} \tag{9}$$

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

$$P_{X_p X_r}(f_k, l) = \beta P_{X_p X_r}(f_k, l-1) + (1-\beta) X_p(f_k, l) X_r^*(f_k, l). \tag{10}$$

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

$$\overline{\mathbf{MSC}}(f_k) = \frac{1}{l} \sum_{i=1}^{l} \left[\rho(f_k, i) \right]^2. \tag{11}$$

The above recursively averaged MSC criterion can thus be used as a means for determining the level of correlation between the disturbing noise sources within the various frequency bands (by averaging the above MSC over each respective linearly or non-linearly spaced sub-band), during the noise alone period in intermittent speech, and consequently selecting the right SBP option, as discussed in section 2.1.

On initial trials, a threshold value around 0.55 for the adaptive MSC has been chosen 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. The above MSC metric was successfully tested for a range of realistic SNR values (from -3dB to 25dB) using both simulated and real reverberant data.

3 Simulation Results

In this section the two new MMSBA-based WF approaches are compared to the original 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 are generated according to the two following schemes.

1. reference noise signal $n(k)$ is chosen to be a random signal, from which two different noise sources (one for primary and one for the reference channel) are derived and summed with $s(k)$ to form x_1, x_2 as in (8).
2. $n_1(k)$, $n_2(k)$ 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 MMSBA (without WF), MMSBA-WF and MMSBA-WBWF. Taking into account the un-correlation between noise and signal on the same channel, we can define the SNR at the output level as:

$$SNR_o(f) = \left[P_{\tilde{Y}\tilde{Y}}(f) - P_{\tilde{N}\tilde{N}}(f) \right] / P_{\tilde{N}\tilde{N}}(f). \tag{12}$$

where the involved power spectra are related to signals described by (3). Similar formulas can be derived considering power spectra in (6) and (7), for MMSBA-WF and MMSBA-WBWF respectively. Moreover it has to be said that $P_{\tilde{N}\tilde{N}}(f)$ 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 different schemes addressed. Choices for various experimental parameter values were selected on a trial and error basis: 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 noise (for both situations addressed); number of iterations of WF operation: 5; number of sub-bands: 4; number of taps or order of FIR adaptive sub-band filters: 32.

Let us consider the results relative to the synthetic noise case study.

- **Coherent Noise:** The intermittent coherent noise-canceller approach is only employed as the SBP option in each band. Table 1 summarizes the results obtained using the three MMSBA approaches: from which it can be seen that MMSBA-WF

and MMSBA-WBWF both deliver an improved SNR performance over the original MMSBA approach.

- **Incoherent Noise:** In this case the value of the recursive MSC metric is used to employ both intermittent and FW SBP options, with the former option used in the first sub-band (with a high MSC) and the latter in the other three bands (with a low MSC). 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 MMSBA-WF scheme) gives the best results in this case, due to its operation in the sub-bands with diverse SBP, resulting in more effective noise cancellation in the frequency domain, compared to the wide-band WF processing (within the MMSBA-WBWF scheme).

Now we can focus our attention to the in-car recorded noise case study.

- **Coherent Noise:** In the first experimental case study, simulated coherent noise over all four bands is used (with a MSC>0.55 in each band), for which the intermittent coherent noise-canceller approach is thus employed as the SBP option in each band. Table 3 summarizes the results obtained using the three MMSBA approaches: from which it can be seen that MMSBA-WF and MMSBA-WBWF both deliver an improved SNR performance over the original MMSBA approach. It is also evident that the choice of sub-band WF (within the MMSBA-WF scheme) gives the best results, as expected, due to its operation in the sub-bands resulting in more effective noise cancellation in the frequency domain, compared to the wide-band WF processing (within the MMSBA-WBWF scheme).
- **Incoherent Noise:** In this more realistic case, simulated incoherent noise over two of the four bands is used. Accordingly in this test case, both intermittent and Ferrara-Widrow SBP options are utilized, the former in the first two sub-bands with highly correlated noises (MSC>0.55 in each band), and the former for the other two bands (with MSC<0.55). Table 4 summarizes the results, from which an even stronger impact of WF operation in the sub-bands (MMSBA-WF processing) is evident.

Note that application of the classical wide-band noise cancellation approach, namely the MMSBA with number of bands set to one and a wideband FIR filter order of 256 (equivalent to product of number of sub-bands and sub-band filter order) was actually found to degrade the speech quality resulting in a negative SNR improvement value, which is hence not shown in the Table 1-4. 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 methods.

4 Conclusions

Two multi-microphone sub-band adaptive speech enhancement systems employing Wiener filtering and a human cochlear model filterbank have been presented. Prelimi-

nary comparative results achieved in simulation experiments demonstrate that the proposed WF based MMSBA schemes are capable of improving the output SNR of speech signals with no additional distortion apparent, compared to the conventional MMSBA scheme (without WF). The MMSBA-WF architecture employing sub-band WF seems to be the most promising whose improved performance is due to the ability of WF to further reduce the residual in-coherent sub-band noise components resulting from MMSBA application. A detailed theoretical analysis is now proposed to define the attainable performance, in addition to employing other perceptive evaluation measures such as the perceptually weighted segmental SNR, Bank Spectral Distortion (BSD) and Perceptual Evaluation of speech Quality (PESQ) scores. 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 various real environments), in order to further assess and quantify the relative advantages of the new speech enhancement schemes.

Table 1. Case A. Synthetic noise. Relative average SNR improvements for all architectures involved (over 10 runs). Standard deviation values are directly depicted on the bars.

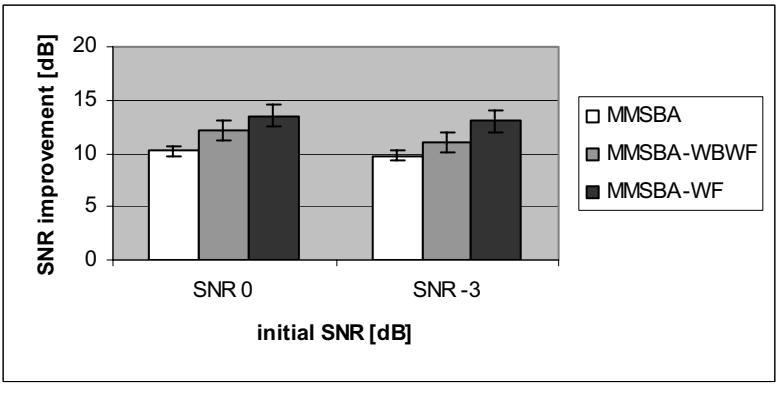


Table 2. Case B. Synthetic noise. Relative average SNR improvements for all architectures involved (over 10 runs). Standard deviation values are directly depicted on the bars.

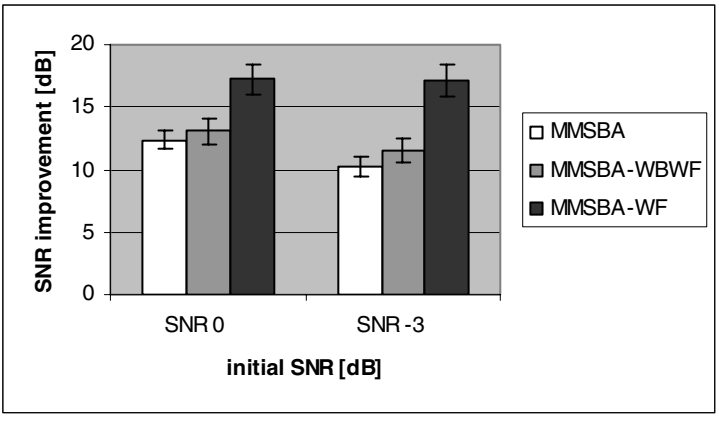


Table 3. Case A. Real in-car noise. Relative average SNR improvements for all architectures involved (over 10 runs). Standard deviation values are directly depicted on the bars.

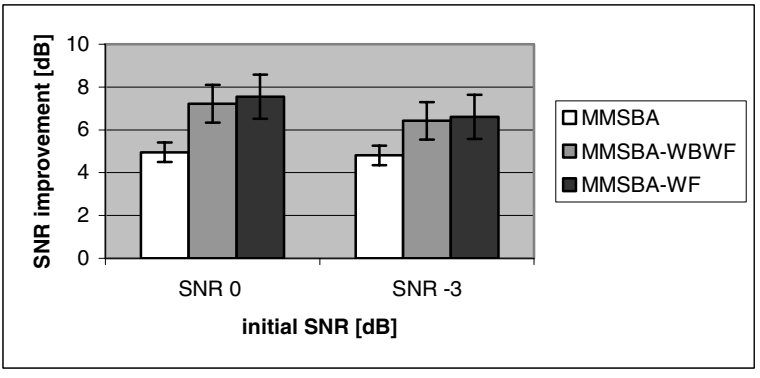
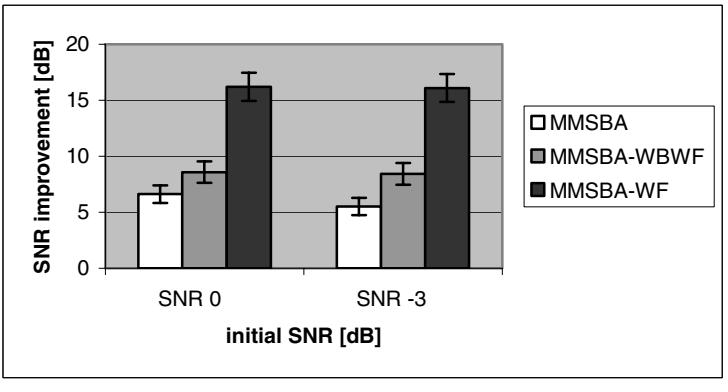


Table 4. Case B. Real in-car noise. Relative average SNR improvements for all architectures involved (over 10 runs). Standard deviation values are directly depicted on the bars.



Finally, further work is currently in progress on employing non-linear sub-band adaptive filtering and cross-band effects (to mimic human lateral inhibition effects) within the binaural MMSBA scheme. These could prove to be more effective in dealing with the non-Gaussian nature of speech and non-linear distortions in the electro-acoustic transmission systems.

References

1. Dabis, H.S., Moir, T.J., Campbell, D.R.: Speech Enhancement by Recursive Estimation of Differential Transfer Functions. Proceedings of ICSP, pp. 345-348, Beijing, 1990
2. Toner, E.: Speech Enhancement using Digital Signal Processing. PhD Thesis, University of Paisley, UK, 1993

3. Darlington, D.J., Campbell, D.R.: Sub-Band Adaptive Filtering Applied to Hearing Aids, Proc.ICSLP'96, pp. 921-924, Philadelphia, USA, 1996
4. Hussain, A., Campbell, D.R.: 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. Abutalebi, H. R., Sheikhzadeh, H., Brennan, R.L., Freeman, G. H.: A Hybrid Sub-Band System for Speech Enhancement in Diffuse Noise Fields. IEEE Sig. Process. Letters, 2003
6. Greenwood, D.D.: A Cochlear Frequency-Position Function for Several Species-29 Years Later. J. Acoustic Soc. Amer., vol. 86, no. 6, pp. 2592-2605, 1990
7. Ferrara, E.R., Widrow, B.: Multi-Channel Adaptive Filtering for Signal Enhancement, IEEE Trans. on Acoustics, Speech and Signal Proc., vol. 29, no. 3, pp. 766-770, 1981
8. Vaseghi, S.V.: Advanced Signal Processing and Digital Noise Reduction (2^{nd} ed.). John Wiley & Sons, 2000
9. Le Bouquin, R., Faucon, G.: Study of a Voice Activity Detector and Its Influence on a Noise Reduction System. Speech Communication, vol. 16, pp. 245-254, 1995