6th European Conference on
Speech Communication and Technology
(EUROSPEECH'99)
Budapest, Hungary, September 5-9, 1999

# INTELLIGIBILITY IMPROVEMENTS USING DIVERSE SUB-BAND PROCESSING APPLIED TO NOISY SPEECH

*Amir Hussain and Douglas R. Campbell**
Department of Applied Computing,
University of Dundee, Dundee, Scotland, DD1 4HN, UK
*Department of Electronic Engineering and Physics,
University of Paisley, Paisley, Scotland, PA1 2BE, UK
E-mail: d.r.campbell@paisley.ac.uk



## ABSTRACT

The paper presents the results of a series of experiments to assess the capability of a multi-microphone sub-band adaptive (MMSBA) signal processing scheme for improving the intelligibility of speech corrupted with noise recorded in an automobile. The noise corrupted speech signals were presented to 15 normal hearing volunteer subjects at various SNRs, and numbers and distributions of sub-bands. The results of listening tests were analysed to determine:

(i) The direction and statistical significance (DSS) of any effect of processing treatment on intelligibility.
(ii) The DSS of any effect of processing treatment on perceived quality.
(iii) The DSS of any effect of the numbers of sub-bands used in sub-band processing (SBP).
(iv) The DSS of any effect of the spacing of sub-bands used in SBP.

In the experimental cases considered, the MMSBA scheme employing diverse sub-band processing is shown to deliver a statistically significant improvement in terms of both speech intelligibility and perceived quality when compared with both the wide-band processed and the noisy unprocessed case.

## 1. INTRODUCTION

A multi-microphone sub-band adaptive (MMSBA) signal processing scheme [1] has shown the potential to yield more than 10dB SNR improvements over conventional full-band methods in real noisy reverberant environments. It can be argued that current research into enhancement of speech corrupted with noise and/or reverberation, places rather too much emphasis on a measure of signal-to-noise ratio (SNR) improvement or speech transmission index; rather than on a quantitative analysis of the improvement in terms of intelligibility [2]. The experiments reported in this paper aim to establish whether MMSBA processing, presented in section 2, delivers a statistically significant and practically useful improvement in intelligibility and/or quality when using speech corrupted with real reverberant noise from an automobile acoustic environment, at SNRs in which normal-hearing individuals would have difficulty.

## 2. THE MMSBA SCHEME

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-alone" period in intermittent speech. Termed the Multi-Microphone sub-band Adaptive (MMSBA) speech enhancement system, the method applies processing within a set of sub-bands provided by a dual-channel filter bank as shown in Figure 1.

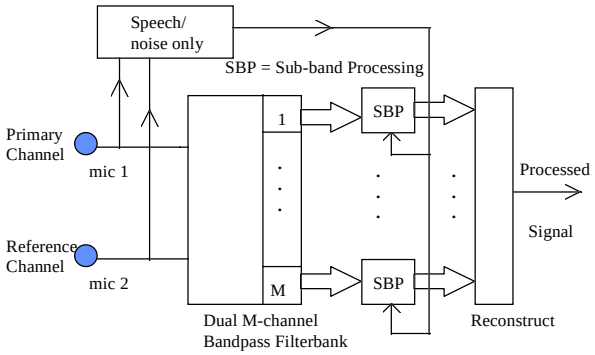


Figure 1: The MMSBA system

The sub-band system decomposes the wide-band input signals into a number of band-limited signals, superficially similar to the treatment the human ear performs on incoming signals. The filter bank can be implemented using various orthogonal transforms or by a parallel filter bank approach. The sub-bands are distributed in either a linear fashion, or a non-linear fashion as in human hearing. The non-linear distribution used here is the cochlear function [3],

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

where $x$ is the proportional distance from 0 to 1 along the cochlear membrane, $A$, $a$ and $k$ are constants based on empirical knowledge of the cochlea, and $F(x)$ are the upper and lower cut-off frequencies for each filter obtained by the limiting value of $x$. For the human cochlea, values of $A$ = 165.4, $a$ = 2.1 and $k$ = 0.88 are suitable [3]. In this work, the division into sub-bands is achieved by modifying the spectra of the FFT (or DCT) of the input signals, and the number of filters is therefore limited by the size of the FFT.

It is assumed in this work that the speaker is close enough to the microphones so that acoustic enclosure

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, and that an effective voice activity detector (VAD) is available.

### 2.2 Diverse Sub-band Processing (SBP) Options

A significant advantage of using sub-band processing (SBP) for speech enhancement is that it allows for diverse processing in each sub-band in order to simultaneously effectively cancel both the coherent and non-coherent noise components present in e.g. automobile environments. The SBP can be accomplished in a number of ways, for example:

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

(ii) *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 [1].

(iii) *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 [4]) may be applied during the noisy speech period as in [1].

In this paper, we employ the above three SBP options and implement the adaptive processing using the Least Mean Squares (LMS) algorithm. In addition we employ a Correlation Metric (CM) (see [1] for mathematical details) based on a recursively estimated Magnitude Squared Coherence (MSC) as part of a system for selecting an appropriate SBP option in the MMSBA scheme.

### 2.3 Correlation Metric (CM) for Selecting SBP

The MSC has been applied to the reduction of noise in speech signals and employed as a VAD for the case of spatially uncorrelated noises [5]. In our work the CM is used as a means for determining the level of correlation between the disturbing noise sources within each frequency sub-band during the "noise-alone" period in intermittent speech. On the basis of this the subsequent form of processing in each respective frequency band can be selected from the SBP options of (i), (ii) and (iii) above.

## 3. HUMAN SUBJECT ASSESSMENT OF SPEECH INTELLIGIBILITY AND QUALITY

### 3.1 Experimental Data Generation

Noise sequences recorded in a Mercedes Benz (1990 Model) were obtained and used in assessing the performance of the MMSBA scheme. The noise sequences were recorded in the car when travelling at 100km/hr on a highway, using two dashboard microphones with a microphone-to-microphone spacing (MTM) of 0.06m and a sampling frequency of 12 kHz. These were added by computer to the 80 different anechoic speech sentences of the Four Alternative Auditory Feature (FAAF) data set [6], to manufacture different easily adjusted SNR cases.

### 3.2 Pilot experiments

A pilot experimental investigation was carried out to determine suitable ranges for the factors to be assessed in the more extensive experiments to follow, and to verify procedures. A subset of the noisy speech data set and three human listeners with verified “normal hearing” were employed in the pilot investigation.

These investigations established a "noise-alone" period of duration approximately 0.9 seconds as giving a consistent estimate of the recursively calculated CM. A suitable CM threshold value for distinguishing between highly correlated and weakly correlated sub-band noise signals was also established. The processing option to be employed within each sub-band for a particular MMSBA scheme was then selected using a combination of the CM, absolute and relative sub-band noise powers and SNR metrics. Based on the pilot trials, two SNR levels (–9dB and –3dB) were selected as eliciting a significant number of errors and being of practical relevance.

The conventional wide-band noise canceller used for comparison was intermittently adapted using the LMS algorithm. The order of the adaptive filter was experimentally set to 1024, and the "noise-alone" period manually labelled to comprise the first 2048 samples (approx. 0.2 seconds). The MMSBA processing schemes to be assessed were selected as operating with 4 & 16 sub-bands for both linear and cochlear sub-band distributions, and used filter lengths which delivered an overall computational complexity similar to that of the wide-band scheme.

### 3.3 Experimental subjects and the subject experience

The pilot experiments were followed by a more substantial set of listening tests involving 6 male and 9 female volunteers between 16 and 45 years age. All subjects had their hearing levels verified as in the normal range by prior audiometric testing. Although all subjects were fluent English speakers, for eight of the subjects it was their second language. Each subject was given a number of clean speech practice sentences until they were familiar with the procedure, and a clean speech score was then established. Subjects were required to listen to FAAF sentences masked by the recorded automobile noise presented at each of the two SNRs, and either unprocessed, processed by a conventional wide-band noise-cancellation approach, or sub-band processed at either of two sub-band spacings and two different numbers of sub-bands. The subjects were presented with the test material audibly via headphones in a single blind, four alternative, forced response protocol. Their task was to identify each of 80 possible target words in a carrier sentence: *‘Can your hear “ *** ”clearly?’* where each target “ *** ” was presented in random order from

one of four alternatives differing by only one phoneme e.g. TIN, BIN, PIN, and DIN. The four possible alternatives were simultaneously presented visually to the subjects as a vertical list on a touch screen monitor. The subject's selection was automatically recorded by the PC based Hearing Assessment Workstation and the score "correct selections out of 80" calculated at the end of each run and stored for later analysis. In this scheme the mean score achieved by chance would be 20/80. The subjects also reported a Mean Opinion Score (MOS) and were instructed to score as follows: Score 1 for reporting "very poor" perceived quality of speech, score 2 for "poor" quality, score 3 for "fair" quality, score 4 for "good" quality and score 5 for "very good" perceived quality of speech.

### 3.4 Experimental variables

The dependent variables were Intelligibility Score as obtained from the FAAF test, and MOS score. The independent variables were: "Subject" having 15 levels (one for each volunteer); "SNR", having two levels (–9dB and –3dB); "Processing", having three levels (Noisy unprocessed, Wide-band processed, Sub-band processed); "NumBands", having two levels (4 and 16 sub-bands); and "Spacing" having two levels (Linear and Cochlear).

### 3.5 Results

Figures 2 and 3 show the mean intelligibility and mean opinion scores (+/- 1 Standard Error) respectively, for the following cases:

*Case 1*: "Clean", Clean speech - no other processing.
*Case 2*: "Noisy Unproc", Noisy speech - no other processing.
*Case 3*: "4 Sub Lin", MMSBA processing scheme with 4-linear spaced sub-bands.
*Case 4*: "16 Sub Lin", MMSBA processing scheme with 16-linear spaced sub-bands.
*Case 5*: "4 Sub Coc", MMSBA processing scheme with 4 cochlear spaced sub-bands.
*Case 6*: "16 Sub Coc", MMSBA processing scheme with 16 cochlear spaced sub-bands.
*Case 7*: "Wide-band", the conventional wide-band noise canceller.
Where cases 2 to 7 each employed the two SNRs, -9dB and -3dB.

One tailed, paired T-tests were applied to the intelligibility and MOS data to assess the statistical significance of the improvement in mean scores evident in the processed schemes (Cases 3 to 7) when compared with unprocessed (Case 2). Two tailed paired T-tests were applied to assess the significance of the differences in scores between the sub-band processing cases.

An analysis of variance (ANOVA) was also performed on the scores using Subject, SNR, Processing, Spacing, and NumBands as factors. Since "Spacing" and "NumBands" are not relevant factors for the noisy speech case, ANOVAs to assess the significance of these factors were applied to the difference between processed and unprocessed score. Table 1 presents a summary of ANOVA results.

| Probability (p) that effects are due to chance | | | |
|---|---|---|---|
| **Factors** | **p Intelligibility** | **p MOS** | **Score** |
| Subject | 0.000 | 0.001 | |
| Processing | 0.000 | 0.000 | Raw |
| SNR | 0.000 | 0.000 | |
| Spacing | 0.012 | 0.000 | Difference |
| NumBands | 0.005 | 0.000 | |

Table 1: ANOVA Summary

### 3.6 Discussion of Intelligibility and MOS results

As is frequently the case in such experiments the volunteer subject was revealed as a statistically significant factor (Table 1).

"SNR" was also shown to be a significant factor (Table 1), as would be expected since we chose its levels on the basis of results from a pilot experiment.

"Processing" was found to be a significant factor at better than the 99% confidence level (Table 1). As can be seen from the plots in Figs 2 and 3, the mean score for wide-band processing was above that for the noisy unprocessed case, and the mean scores for all forms of sub-band processing were above that for wide-band processing. Taking the ANOVA results together with the results of the T-tests we conclude that for the experimental cases considered, diverse sub-band processing delivers a statistically significant improvement in terms of both speech intelligibility and perceived quality when compared with both wide-band processed and the noisy unprocessed case.

"NumBands" is shown to be a significant factor (Table 1) at better than the 99% confidence level. "Spacing" is also shown to be a significant factor (Table 1) at better than the 95% confidence level. Relative to the noisy unprocessed case, the 16 linear sub-bands case delivered the maximum improvement in mean intelligibility score (+9.7 @ -3dB SNR, +11.7 @ -9dB SNR). The corresponding increase in intelligibility was calculated to be roughly equivalent to a 2dB to 3dB SNR improvement under the conditions of Foster and Haggard [6], and is a practically useful improvement [7]. This conclusion is supported by the improvement in MOS score (+1.5@ -3dB SNR, +1.5 @ -9dB SNR) for the 16 linear sub-bands case compared with the noisy unprocessed case.

Across all the ANOVAs the major interaction was between "Subject" and "SNR" ($p <= 0.083$). Only in one case did another interaction pair exceed this and that was "Subject" and "Spacing" ($p = 0.047$) for the ANOVA applied to MOS data for the assessment of the effect of sub-band spacing. This implies that spacing has a more demonstrable effect on perceived quality of processed speech, than on intelligibility. In no case did sub-band processing result in a deterioration of intelligibility score or MOS. The correlation was computed between mean intelligibility scores (MIS) and MOS at both SNRs. As reflected in Figs 2 and 3, a strong correlation appeared between the results of the intelligibility tests and the

MOS tests, implying consistency in judgements made by the subjects.

## 4. CONCLUSIONS

The listening test results show that the MMSBA scheme can improve the intelligibility and quality of speech corrupted with real automobile noise. In the scenarios considered, diverse SBP delivers a statistically significant improvement in terms of both speech intelligibility and perceived quality when compared with both the wide-band processed and the noisy unprocessed cases. Both the number of sub-bands and their distribution were found to be significant factors. There is some evidence that sub-band spacing has a more demonstrable effect on perceived quality of processed speech than on intelligibility. In no case did sub-band processing result in a deterioration of intelligibility score or MOS.

## 5. ACKNOWLEDGEMENTS

The U.K. Engineering and Physical Sciences Research Council (EPSRC) Project Reference GR/K48907 supported this work. The authors are grateful to: Dr. K Linhard of Daimler-Benz Ltd., Germany, for providing the automobile noise data; Prof S Gatehouse of The Institute for Hearing Research (IHR), Glasgow Royal Infirmary for supplying the FAAF speech data; Our colleague, Mr. P Shields, contributed to the statistical analysis.

## 6. REFERENCES

[1] A.Hussain, D.R.Campbell and T.J.Moir, *Diverse processing in cochlear-spaced sub-bands for …*, Signal Process. IX, Theor.&Appl., ,**3**, 1489-1492,1998.
[2] C. Ludvigsen, C.Elberling, G.Keidser, *Evaluation of a Noise reduction Method- Comparison …* Scandinavian Audiology, Suppl. **38**, 50-55, 1993.
[3] 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.
[4] E. R. Ferrara, B. Widrow, *Multi-channel Adaptive Filtering for signal enhancement*, IEEE Trans. ASSP., **29**(3),766-770, 1981.
[5] R.L.Bouquin, G.Faucon, *Study of a voice activity detector and its influence on a noise reduction system*, Speech Communication, **16**, 245-254, 1995.
[6] J.R. Foster, M.P.Haggard, *The Four Alternative Auditory Feature (FAAF) test-…*, Jour. British Audio., **21**, 165-174, 1987.
[7] R. Plomp, "*Noise, amplification, and compression: Considerations of three main issues in hearing aid design*", Ear Hear., **15**, 2-12, 1994.

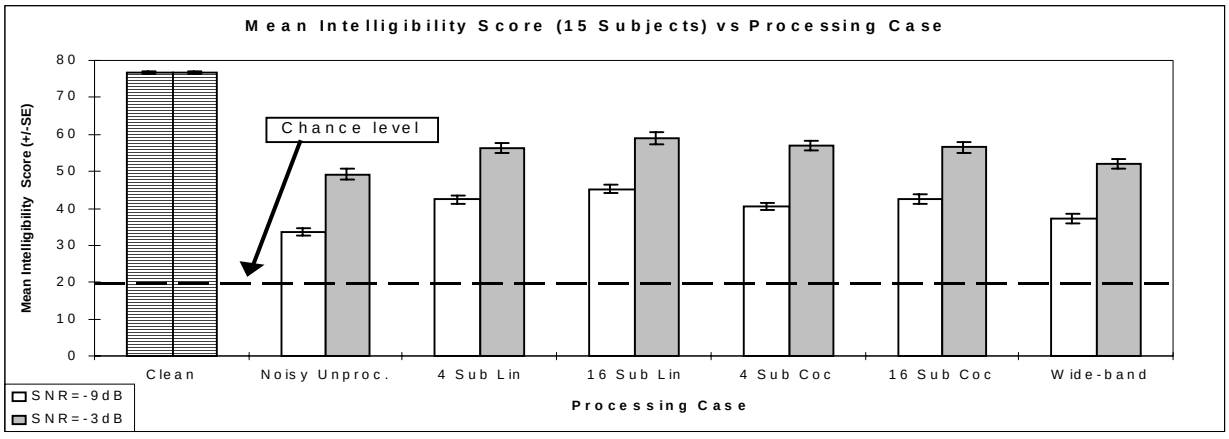


Figure 2

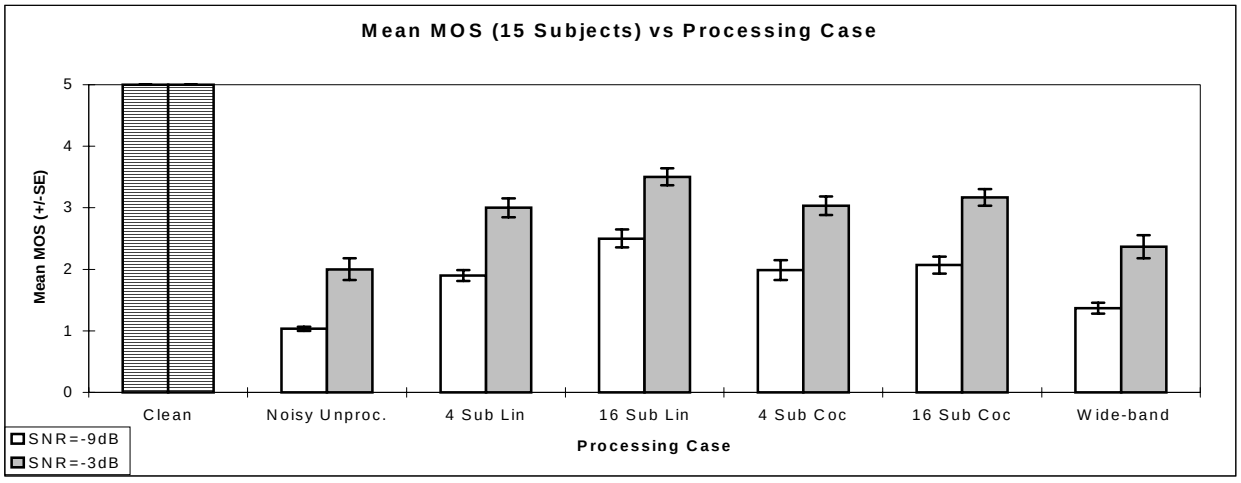


Figure 3