SPEECH ENHANCEMENT USING NEAR-FIELD SUPERDIRECTIVITY WITH AN ADAPTIVE SIDELOBE CANCELER AND POST-FILTER

Iain A. McCowan Darren Moore Sridha Sridharan
Speech Research Laboratory, RCSAVT, School of EESE
Queensland University of Technology
GPO Box 2434, Brisbane QLD 4001, Australia
email: [i.mccowan,dc.moore,s.sridharan]@qut.edu.au

ABSTRACT: This paper describes a new microphone array technique and investigates its effectiveness for speech enhancement. A system structure consisting of a fixed near-field superdirective beamformer and an adaptive sidelobe canceling path is proposed (NFSD-ASC). The effect of adding a post-filter is also examined. The system is evaluated in terms of speech quality measures in the context of a computer workstation in an office environment. The speaker is located directly in front of the computer monitor at a distance of 60 cm and the array is designed to fit across the top of a standard 17 inch monitor. The experiments show that the array is effective in both decreasing the noise level and the amount of signal distortion when compared with standard near-field superdirectivity and the generalised sidelobe canceler.

INTRODUCTION

With speech processing techniques being increasingly applied in real noise environments, speech enhancement is currently an important area of research. Many different enhancement approaches exist, each taking into account different types of knowledge about the desired signal, such as speech production models or spectral content. In this paper, we focus on the use of spatial information to enhance the desired signal by using a microphone array. Microphone array techniques allow spatial filtering of the noisy input, enhancing speech from the desired direction while attenuating noise from all other directions.

It is well known that standard delay-sum beamforming is not well suited to the task of speech enhancement, and many other techniques have been proposed. Among these, superdirective techniques have shown promising performance for practically sized microphone arrays (Doerbecker, 1997; Bitzer et al., 1999). In particular, near-field superdirectivity (Täger, 1998) is well suited to enhancement in the situation where the desired speaker is located in the array's near-field.

The current research seeks to build upon standard near-field superdirectivity by incorporating an adaptive noise canceling path in the system, similar to the generalised sidelobe canceler structure (Griffiths and Jim, 1982). The adaptive noise canceler enables the array to adapt to varying noise conditions, providing additional attenuation to undesired noise sources and leading to lower noise power in the beamformed output. The use of a suitable post-filter to further enhance the desired signal is also investigated.

MICROPHONE ARRAY ENHANCEMENT SYSTEM

Microphone Array

The microphone array is the 11 element array shown in Figure 1. It consists of a 9 element broadside array, with an additional 2 microphones situated directly behind the end microphones. The total array is 40 cm wide and 15 cm deep, lying in the horizontal plane. The broadside microphones are arranged according to a standard broadband sub-array design, where different sub-arrays are used for different frequency ranges. The two endfire microphones are included for use in the low frequency range where the amplitude difference between sensors is greater and can be exploited by the NFSD algorithm. The four sub-arrays are thus as follows :

- $(f < 1\,kHz)$: microphones 1-11;

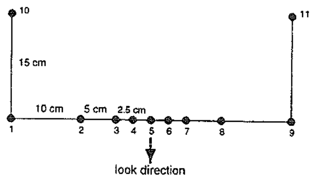

Figure 1: Array Geometry

- $(1\,kHz < f < 2\,kHz)$: microphones 1, 2, 5, 8 and 9;
- $(2\,kHz < f < 4\,kHz)$: microphones 2, 3, 5, 7 and 8; and
- $(4\,kHz < f < 8\,kHz)$: microphones 3-7.

Beamforming Technique

Near-field superdirectivity (Täger, 1998) is a beamforming technique that addresses the problem of poor low frequency performance by compensating for both amplitude and phase differences. The principal assumption made is that the desired source is situated in the array's near-field while the dominating noise sources are located in the far-field, as is generally the case in the chosen application. This assumption is used in the formulation of a near-field modified expression for the factor of directivity. The NFSD algorithm is then essentially a constrained optimisation problem, which seeks to calculate a set of linear array filters which maximise the directivity factor, under

1. a linear constraint of non-distortion of the desired signal;
2. a non-linear constraint of a chosen minimum value for the incoherent noise reduction at each frequency, $G_{a\,min}(f)$; and
3. a set of additional linear constraints (zeros in given directions, main lobe width, etc).

The solution is obtained using the Lagrange method, giving the optimal channel filters $b_i(f)(i = 1, \ldots, N)$ (where N is the number of microphones in the array). The technique is fully described in Täger (1998).

While NFSD has been shown to be quite successful in enhancing the desired signal, its formulation takes no account of the location or frequency characteristics of the noise sources. We therefore seek to address this limitation by using knowledge of the desired signal's direction to form a dynamic estimate of the noise emanating from all other directions. By subtracting this noise estimate from the NFSD beamformed output, we expect to further reduce the noise level in the enhanced output signal.

To incorporate such a noise-canceling path into the system, we use the adaptive portion of the generalised sidelobe canceler (GSC) structure (Griffiths and Jim, 1982). GSC is a beamforming structure that can be used to implement a variety of linearly constrained adaptive array processors. It separates the adaptive beamformer into two main processing paths - a standard fixed beamformer with constraints on the desired signal, and an adaptive path, which consists of a blocking matrix and a set of FIR filters that adapt in order to minimise the power at the ouput. The blocking matrix is effectively a bank of narrowband beamformers, each having a beampattern that has a null in the direction of the desired signal. Its purpose is to exclude the desired signal from the adaptive path, thus ensuring that the output power minimisation process does not distort the desired signal. Use of the adaptive sidelobe canceling portion of the GSC in conjunction with a fixed NFSD beamformer will thus result in lower noise power than that obtained using standard NFSD.

A block diagram of the resulting NFSD-ASC beamformer is given in Figure 2. It is assumed that the inputs have undergone time delay compensation to align the desired signal.

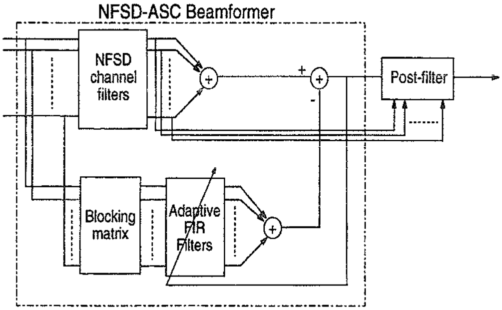

Figure 2: NFSD-ASC Beamformer with Post-filter

Post-filter

A further extension to the beamformer is the addition of a post-filter to the system. A thorough study of the use of post-filters with a filter-sum microphone array was undertaken by Marro et al. (1998), who showed theoretically how such a post-filter can enhance the performance of the beamformer. The availability of multi-channel input allows the Wiener filter transfer function to be estimated using a combination of the auto- and cross-power spectral densities of the filtered input channels. Taking into account the NFSD channel filters, we estimate the post-filter transfer function as

$$\hat{W}(f) = \frac{\sum_{i=1}^{N} |b_i(f)|^2}{\Re\left\{\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} b_i(f)b_j^*(f)\right\}} \times \frac{\Re\left\{\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} \hat{\Phi}_{v_iv_j}(f)\right\}}{\sum_{i=1}^{N} \hat{\Phi}_{v_iv_i}(f)} \quad (1)$$

where v_i is the i^{th} microphone signal filtered by the array filter $b_i(f)$. The values $\hat{\Phi}_{v_iv_j}(f)$,(respectively $\hat{\Phi}_{v_iv_i}(f)$), are the estimated cross (power) spectral densities of signals v_i and v_j. The post-filter provides an additional level of noise adaptation by shaping the output frequency spectrum to enhance the spectral features of the signal and attenuate those of the noise. This differs from the approach taken by the adaptive sidelobe canceler, which subtracts a noise estimate from the output signal. Previous work has shown that optimal use of such a post-filter with a standard NFSD beamformer improves the speech recognition performance in the given noise environment (McCowan et al., 2000).

EXPERIMENTS AND RESULTS

Experimental Configuration

The context of the experiments is the computer room shown in Figure 3. Two different sound source locations were used, these being :

1. the desired speaker situated 60 cm from the centre microphone, directly in front of the array; and
2. a localised noise source at an angle of 56 degrees and a distance of 2.7 metres from the array.

Impulse responses of the acoustic path between each source and microphone were measured from recordings made in the room with the array. The multi-channel desired speech and localised noise microphone inputs were then generated by convolving the speech and noise signals with these impulse responses. In addition, a real multi-channel background noise recording of normal operating conditions was made in the room with other workers present. This recording is referred to in the experiments as the ambient noise signal. It consists mainly of computer noise, a variable level of background speech, and noise from an air-conditioning unit.

Speech Quality Assessment

For the desired speech signal, we randomly chose a segment of speech from the TIDIGITS database corresponding to the digit sequence *one-nine-eight-six*. This was added to the ambient noise recording

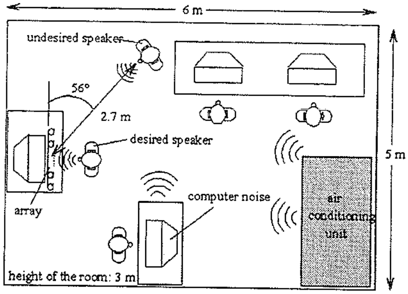

Figure 3: Experimental Setup

technique	*SLNR (dB)*					
	20	*15*	*10*	*5*	*0*	*-5*
noisy input	9.4	8.7	7.3	4.9	2.5	0.8
delay-sum	8.6	8.1	6.9	4.9	2.5	0.7
NFSD	10.3	9.8	8.6	6.4	3.6	1.3
NFSD-PF	11.5	11.0	9.8	7.5	4.6	2.1
GSC	12.6	11.6	10.3	8.5	6.3	4.0
NFSD-ASC	14.9	14.1	12.8	10.7	8.0	5.2
NFSD-ASC-PF	16.7	15.8	14.2	11.9	8.9	5.6

Table 1: Signal to Noise Ratio : localised speech noise and ambient noise

at an average segmental SNR level of 10 dB. In addition, a localised white noise signal was added, corresponding to the location of the undesired speaker in the diagram. The white noise was added at a level of approximately 0 dB. The resulting signals are shown graphically in Figure 4. These show that NFSD-ASC succeeds in reducing the noise level while introducing negligible distortion to the desired signal.

A set of experiments was conducted in which the localised white noise source was replaced with a localised speech-like noise source. The ambient noise recording was still present at an SNR level of 10 dB. The level of the localised speech noise source was varied, and the signal to noise ratios and log area ratios for the enhanced output were calculated. Tables 1 and 2 give the resulting objective quality measures for the different enhancement techniques. Note that the levels in the table headings indicate the signal to localised noise ratio (SLNR), without taking account of the additional ambient noise, whereas the results are SNR values that include all noise sources. These results are plotted in Figure 5.

technique	*SLNR (dB)*					
	20	*15*	*10*	*5*	*0*	*-5*
noisy input	3.0	3.1	3.3	3.6	4.0	4.2
delay-sum	2.9	2.9	3.1	3.5	4.0	4.4
NFSD	2.9	3.0	3.2	3.5	4.0	4.4
NFSD-PF	3.4	3.5	3.6	3.9	4.4	5.0
GSC	2.3	2.4	2.6	3.0	3.5	4.0
NFSD-ASC	2.2	2.3	2.5	2.9	3.4	3.9
NFSD-ASC-PF	2.2	2.3	2.4	2.5	2.8	3.4

Table 2: Log Area Ratio : localised speech noise and ambient noise

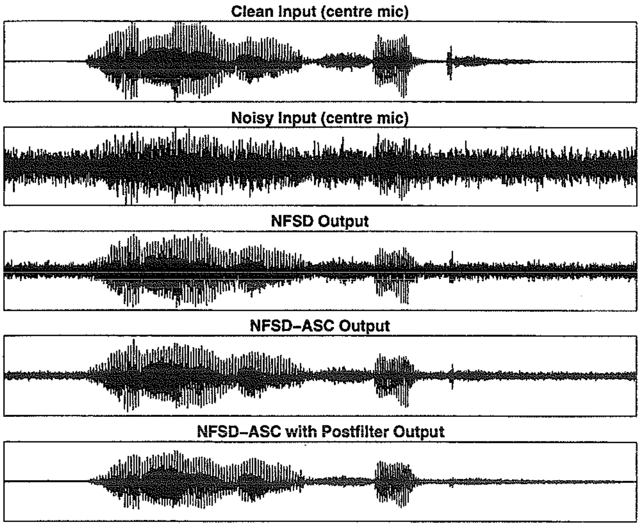

Figure 4: Sample Enhanced Signal

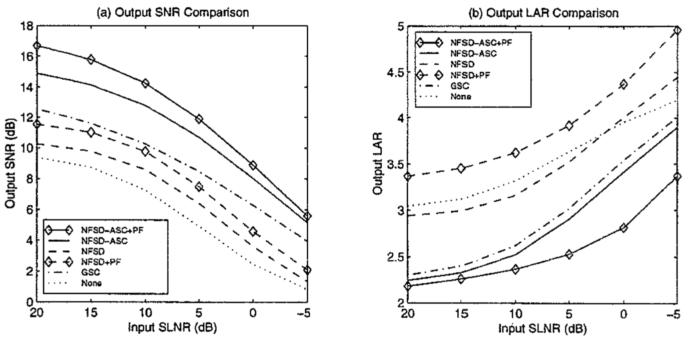

Figure 5: Speech Quality Comparison : (a) SNR and (b) LAR

The results show that the proposed NFSD-ASC is consistently superior to standard GSC, NFSD and NFSD with the post-filter, in terms of both noise reduction and perceptual distortion. This is clearly visible in the plots of the results in Figure 5. In terms of SNR, the proposed NFSD-ASC technique consistently provides an improvement of 5 dB over standard NFSD, and 2 dB over standard GSC. The addition of the post-filter gives an additional 3.5 dB of improvement. Even in highly adverse noise conditions, the technique is successful in reducing the noise level with respect to the desired signal.

While SNR is a valuable indication of noise reduction, the log area ratio (LAR) is known to give a better indication of perceptual intelligibility (Quackenbush et al., 1988). The LAR is an objective measure of the dissimilarity between the linear predictive coefficients of the original and processed speech signals. It is seen that the NFSD-ASC and NFSD-ASC-PF give less distortion than all other techniques assessed. This shows that the noise reduction is achieved without sacrificing the quality or intelligibility of the desired signal, indicating the success of the adaptive sidelobe cancelling path. Thus we see that while the NFSD technique offers a similar level of distortion as a standard delay-sum beamformer, NFSD-ASC succeeds in significantly reducing the perceptual distortion in the enhanced signal.

The addition of the post-filter to NFSD-ASC gives considerable further improvement in terms of both the SNR and LAR measures. It is interesting to note that the LAR distortion increases with the addition of the post-filter to NFSD, but decreases slightly for the new NFSD-ASC technique.

CONCLUSIONS

This paper has introduced a new beamforming technique that combines near-field superdirective beamforming with an adaptive sidelobe canceler. The new technique is clearly shown to give superior performance to both NFSD and standard GSC in the experiments which examine the case of a high degree of office-type noise and a localised speech noise source. Most importantly, as well as providing significant noise reduction, the proposed NFSD-ASC technique is successful in decreasing the perceptual distortion, as measured by the log area ratio, when compared with other beamforming techniques. The addition of a suitable post-filter has also been shown to further reduce both the noise level and the perceptual distortion in the enhanced output.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the assistance of Claude Marro from France Telecom R&D with the research presented in this paper.

REFERENCES

Bitzer, J., Simmer, K. U. and Kammeyer, K. (1999), Multi-microphone noise reduction by post-filter and superdirective beamformer, *Proceedings of IWAENC-99*, pp. 27–30.

Doerbecker, M. (1997), Speech enhancement using small microphone arrays with optimized directivity, *Proc. Int. Workshop on Acoustic Echo and Noise Control*, pp. 100–103.

Griffiths, L. and Jim, C. (1982), An alternative approach to linearly constrained adaptive beamforming, *IEEE Trans. on Antennas and Propagation* **30(1)**, 27–34.

Marro, C., Mahieux, Y. and Uwe Simmer, K. (1998), Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering, *IEEE Transactions on Speech and Audio Processing* **6**(3), 240–259.

McCowan, I., Marro, C. and Mauuary, L. (2000), Robust speech recognition using near-field superdirective beamforming with post-filtering, *Proceedings of ICASSP 2000*, Vol. 3, pp. 1723–1726.

Quackenbush, S., Barnwell, T. and Clements, M. (1988), *Objective Measures of Speech Quality*, Prentice-Hall, NJ.

Täger, W. (1998), Near field superdirectivity (NFSD), *Proceedings of ICASSP '98*, pp. 2045–2048.