A Study of a various Acoustic Beamforming Techniques Using a Microphone Array

Ahmed Abdalla*, Peiyu He
School of Electronics and Information Engineering, Sichuan University
Chengdu, China
Email: *ahmed.baoney6@hotmail.com

***Abstract*— In this paper, A study of numerous acoustic beamforming algorithms is carried out. Beamforming algorithms are techniques utilize to determine the Direction of arrival of (DOA) the speech signals while suppress out the corresponding noises and interferences. The simple delay and sum beamformer technique which use the constrained least mean squares (LMS) filter for spatial filtering is firstly investigated. Secondly, a constrained least mean square algorithm (also known as Frost Beamformer) is considered. The beamformer algorithms are simulated in MATLAB and it observed that there a significant enhancement in the Signal-to-Noise-Ratio (SNR) for frost beamformer as compared to the simple delay and sum beamformer.**

Keywords—microphones linear array; Frost Beamformer; delay and sum beamformer; speech signal.

I. Introduction

Mainly speech signal receiving systems work in noisy environments, where the desired speech signal is corrupted by interfering signals, and also distorted by the reverberating environment. In many applications, there is a need to separate the multiple sources or extract a source of interest while minimizing undesired interfering signals and noise [1]. It is confirmed scientifically that humans are capable to separate only one conversation in extremely noisy environments, such as in a cocktail party environment. Notwithstanding of being studied for decades, the cocktail party problem remains a scientific challenge that stress further research efforts [2].

Obviously, in some recent works [4], using a single channel it is not achievable to improve both clearness and quality of the recovered signal at the same time. A method to overcome this drawback is to add some spatial information to the time/frequency information obtainable in the single channel case. Furthermore, one can get this additional information with two or more channels of noisy speech known as multichannel. It should be noted that there are two main categories of multichannel (microphone array) algorithms, namely: Blind Source Separation (BSS) and Beamforming. BSS is an approach for determining source signals using only information about their mixtures observed in each input channel. The estimation is performed without possessing information on each source, such as its frequency characteristics and location, or on how the sources are mixed. On the other hand, the beamforming families of algorithms concentrate on enhancing the sum of the desired sources while treating all other signals as interfering sources.

A Beamformer is a signal processor used jointly with a microphone array to provide the capability of spatial filtering. The microphone array produces spatial samples of the propagating wave, which are then manipulated by the signal processor to make the beamformer output signal. Beamforming is achieved by filtering the microphone signals and combining the outputs to extract (by constructive combining) the desired signal and reject (by destructive combining) interfering signals according to their spatial location. Beamforming can split sources with overlapping frequency content that originate at different spatial locations. Speech separation based on beamforming techniques have been intensively considered in latest years due to their many applications. These techniques can be divided into two categories, depending on the approach in use to estimate the spatial filter weights: deterministic and statistically beamforming approaches.

This paper aims to study the delay and sum beamformer and the frost beamformer. Sufficiently, these methods are highlighted and compared under different viewpoints. Furthermore, the simulations are implemented to confirm the theoretical analysis.

The remainder of the paper is organized as follows. The next section is devoted to beamformers classification while Section III contains delay and sum beamformer definition and investigation while section IV examines the frost beamformer. In Section V, we report the adaptive algorithms of the Frost beamformer. In Section VI the simulations are illustrated in order to evaluate the performance of the methods. Finally, concluding remarks are given in Section VII.

II. Beamformers Classification

Beamformers are classified as either data independent or statistically optimum, depending on how the weights are selected. The weights in a data independent beamformer do not depend on the array data and are chosen to present a specified response for all signal and interference scenarios. The weights in a statistically optimal beamformer are chosen based on the statistics of the array data to optimize the array response [6]. Meanwhile, the statistics of the array data are not usually known and may vary over time, so adaptive algorithms are normally used to determine the weights. Therefore, the adaptive algorithm is designed and hence the beamformer response converges to a statistically optimum solution.

Conversely, the weights in a data independent beamformer are designed so that the beamformer response approximates a preferred reaction independent of the array data or data statistics. Apparently these propose objective is identical as that for a classical FIR filter design. A straightforward the simple Delay and sum beamformer is an example of the data independent beamforming.

Consequently, in statistically optimum beamformer the weighs are selected base on the statistics of the data received at the array. The aim is to optimize the beamformer response and thus the output signal includes smallest contributions due to the noise and signals arriving from directions other than the desired direction. Evidently, the frost beamformer is a statistically optimum beamformer. Moreover, other statistically optimum beamformers are multiple side lobe canceller (MSLC) and maximization of the signal to noise ratio.

III. DELAY AND SUM BEAMFORMER

Generally, the underlying idea of sum-and-delay beamforming is that when an electromagnetic signal impinges upon the aperture of the antenna array, the element outputs, added together with suitable amounts of delays, reinforce signals with regard to noise or signals arriving at unlike directions. We should point out that, the delays requested rely on the physical spacing between the elements in the array. The geometrical arrangement of elements and weights coupled with each element are crucial factors in identifying the array's characteristics.

Predominately in delay-and-sum beamforming, delays are positioned after each microphone to compensate the arrival time differences of the speech signal to each microphone as can be seen in Fig. 1. Ultimately the time aligned signals at the outputs of the delays are then summed together. Thus, this has the result of reinforcing the desired speech signal while the unwanted off-axis noise signals are mixed in a more unforeseeable style. Due to this fact, the signal-to-noise ratio (SNR) of the total signal is henceforth greater than or at worst, equal to that of any unique microphone’s signal. Accordingly, this system makes the array pattern further susceptible to sources from a particular desired direction.

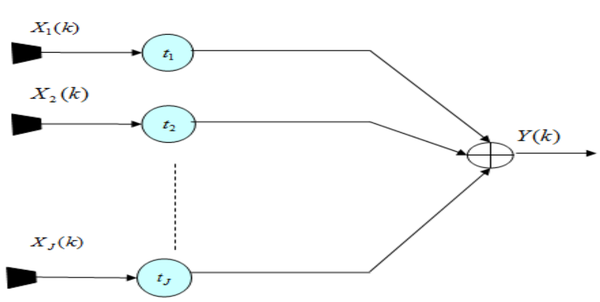

Fig. 1. Delay and Sum Beamformer with J sensors.

The major disadvantage of delay-and-sum beamforming systems is the large number of sensors required to improve the SNR, which it is inflexible in application, since it may lead to higher system complexity and more additional equipment. Each doubling of the number of sensors will afford at most an additional 3 dB increase in SNR, yet this is if and only if the incoming jamming signals are completely uncorrelated between the sensors and with the desired signal. Furthermore, another disadvantage is that no nulls are located directly in jamming signal locations. Hereby the delay-and-sum beamformer attempts only to improve the signal in the direction to which the array is currently steered.

IV. FROSST BEAMFORMER

The Constrained Least Mean Squares or Constrained LMS algorithm is a straightforward stochastic gradient algorithm which needs only the DOA and the desired frequency response in the look direction. Nerveless, in the adaptive process, the algorithm gradually become skilled at statistics of noise arriving from directions different than look direction. Moreover, the algorithm is able to preserve a chosen frequency response in the look direction while minimizing output noise power. Consider the array processor shown in Figure 2. Obviously the processor has K sensors and J taps per sensor. Therefore, there are KJ weights. Out of these J weights decide the look direction frequency response [7].

In the Fig. 2 the delays after each sensor are not shown. The array processor is supposed to be steered to the required look direction by suitable delays after the sensors the same as in the case of delay and sum beamforming. The remaining KJ – J weights may be utilized to reduce the total power in the array output. It should be emphasized that, minimization of the total output power is equivalent to minimizing the non-look direction noise power as long as the signal and the noise is uncorrelated which is a reasonable statement.

As well as the signal is concerned, the array processor is equivalent to a single tapped delay in which each weight is identical to the sum of the weights in the vertical column of the processor. These summation weights in the equivalent tapped delay line necessity to be selected since are providing the desired frequency response characteristic in the look direction.

The vector of tap voltages at the k-th sample can be written as $X(k)$ where

$$X^T(k) = \left[x_1(k), x_2(k), \ldots\ldots\ldots\ldots, x_{KJ}(k)\right] \qquad (1)$$

The tap voltages are the sums of the voltages due to look-direction waveforms and the non-look-direction noises, so that

$$X(k) = L(k) + N(k) \qquad (2)$$

Where the KJ dimensional vector of look-direction at the kth sample can be recast as

$$L^T(k) = [\underbrace{l(k) \;\; . \;\; . \;\; l(k)}_{\text{k taps}} \;\; \underbrace{l(k-1) \;\; . \;\; . \;\; l(k-1)}_{\text{k taps}} \;\; . \;\; . \;\; \underbrace{l(k-(J-1)) \;\; . \;\; . \;\; l(k-(J-1))}_{\text{k taps}}]$$

And the vector of non-look-direction noises can be shown as

$$N^T(k) = \left[n_1(k) \quad n_2(k) \quad . \quad . \quad . \quad n_{KJ}(k)\right] \qquad (3)$$

The vector of weights at each tap is W, where

$$W^T = [w_1 \quad w_2 \quad . \quad . \quad . \quad w_{KJ}] \tag{4}$$

$$\begin{aligned} E\left[X(k)X^T(k)\right] &= R_{XX} \\ E\left[N(k)N^T(k)\right] &= R_{NN} \\ E\left[L(k)L^T(k)\right] &= R_{LL} \end{aligned} \tag{5}$$

Meanwhile it is assumed that the look direction waveform is uncorrelated with the vector of non-look direction noise, thus the following consideration must take in account which can be expressed as

$$E\left[N(k)L^T(k)\right] = 0 \tag{6}$$

Ultimately, the output of the array at the time of the kth sample is

$$y(k) = W^T X(k) = X^T(k)W \tag{7}$$

The constraints that the weights on the jth vertical column of the taps sum to a chosen number f_i is expressed by the requirement

$$\begin{aligned} & c_j^T W = f_i \qquad j = 1,2,......,J \\ & \min_{W} \; W^T R_{XX} W \\ & \textit{subject to } \; C^T W \end{aligned} \tag{8}$$

The aforemetional eqytion is the constrained LMS problem. W_{opt} Is establish by the method of Lagrange multipliers and then

$$H(W) = \frac{1}{2} W^T R_{XX} W + \lambda^T (C^T W - F) \tag{9}$$

Taking the gradient with respect to W yield $\nabla_W H(W) = R_{XX} W + C\lambda$ and then setting this to zero the result can be recast as

$$W_{opt} = -R_{XX}^{-1} C \lambda \tag{10}$$

Since R_{XX} is positive semi definite if the inverse exists. Substituting this in the constraint equation (8) can be formulated as bellow

$$C^T W_{opt} = F = -C^T R_{XX}^{-1} C\lambda \tag{11}$$

Apparently the Lagrange multipliers can be recast as $\lambda = -\left[C^T R_{XX}^{-1} C\right]^{-1} F$. Therefore, the best weight vector can be written as

$$W_{opt} = R_{XX}^{-1} C \left[C^T R_{XX}^{-1} C\right]^{-1} F \tag{12}$$

V. Adaptive Algorithms Of The Frost Beamformer

In order to find the optimum weights, the input correlation matrix R_{XX} is not known a priori and must be well learnt by an adaptive technique. Direct substitution of a correlation matrix estimate into the optimal weight equation necessitated a number of multiplications at each iteration commensurate to the cube of the number of weights. Certainly, the complexity is due to the inversion of the input correlation matrix. The adaptive algorithm take in consideration requires only a number of multiplications and storage locations directly proportional to the number of weights. In constrained gradient-descent optimization, the weight vector is initialized at a vector satisfying the constraint say$W(0) = C(C^T C)^{-1} F$, and at each iteration the weight vector is moved in the negative direction of the constrained gradient. The length of the step is directly proportional to the amplitude of the constrained gradient and is scaled by a constant μ . Thus,after the kth iteration the next weight vector can be recast as

$$\begin{aligned} W(k+1) &= W(k) - \mu \nabla_W H[W(k)] \\ &= W(k) - \mu[R_{XX} W(k) + C\lambda(k)] \end{aligned} \tag{13}$$

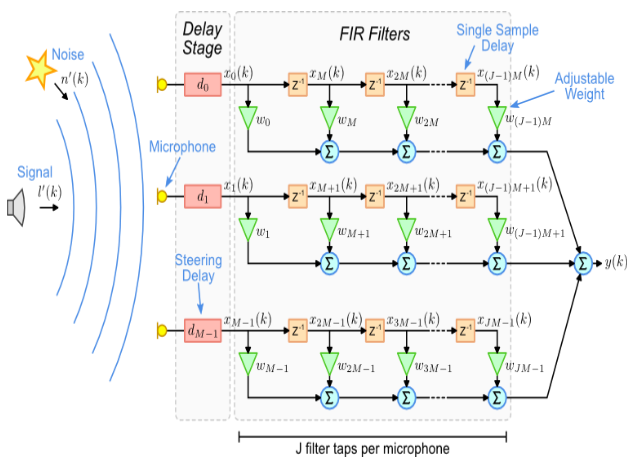

Fig. 2. Frost Beamformer.

The Lagrange multipliers are chosen by requiring W (k+1) to satisfy the constraint

$$\begin{aligned} F &= C^T W(k+1) = C^T W(k) - \mu C^T R_{XX} W(k) \\ &\quad - \mu C^T C \lambda(k) \end{aligned} \tag{14}$$

Solving for the Lagrange multipliers λ(k) and substituting into the weight-iteration equation (13) can be recast as

$$\begin{aligned} W(k+1) &= W(k) - \mu\left[I - C(C^T C)^{-1} C^T\right] R_{XX} W(k) \\ &\quad + C(C^T C)^{-1}\left[F - C^T W(k)\right] \end{aligned} \tag{15}$$

Defining the KJ dimensional vector $\tilde{F} = C(C^T C)^{-1} F$ and the and the KJ x KJ matrix $P = I - C(C^T C)^{-1} C^T$ the algorithm may be written as bellow

$$W(k+1)=P\left[W(k)-\mu R_{XX}W(k)\right]+\tilde{F} \tag{17}$$

A straightforward approximation for R_{XX} at the kth iteration is the outer product of the tap voltage vector with itself: henceforth the stochastic constrained LMS algorithm is $W(0)=\tilde{F}$ thus (17) can be finally calculate by

$$W(k+1)=P\left[W(k)-\mu y(k)X(k)\right]+\tilde{F} \tag{18}$$

VI. Simulations

The time delay beamformer and the frost beamformer had 10 omnidirectional microphones (sensors) placed in a linear array with the distance between the sensors d=0.5 m. For the Adaptive frost beamformer each sensor branch had 20 taps. The environment consisted of two recorded speeches and one laughter recording. We also load the laughter audio segment as interference and white noise signal with a power of 1e-4 watts to represent the thermal noise. The sampling frequency of the audio signals is 8 kHz. The incident direction of the first speech signal is -30 degrees in azimuth and 0 degrees in elevation. The direction of the second speech signal is -10 degrees in azimuth and 10 degrees in elevation. The interference arrives from 20 degrees in azimuth and 0 degrees in elevation.

At first, the output results of the received signals before and after suppressing the interference of delay and sum and frost beamformer methods are given to illustrate the effectiveness of these methods.

Fig. 3 shows the output signals of two desired signals and one interference signal on any channel. Whereas Fig. 4. Shows output after process the signal by time delay beamformer technique, and it is clear that the methods improve the SNR. Fig. 5 shows the output of the received signal after use the frost beamformer method

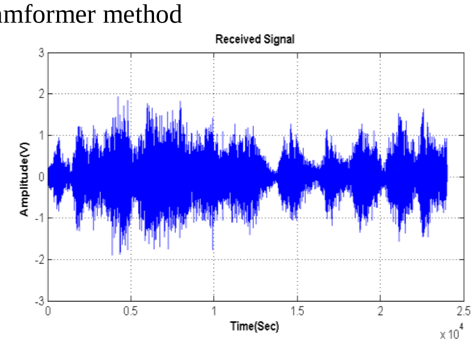

Fig. 3. The Received Signal.

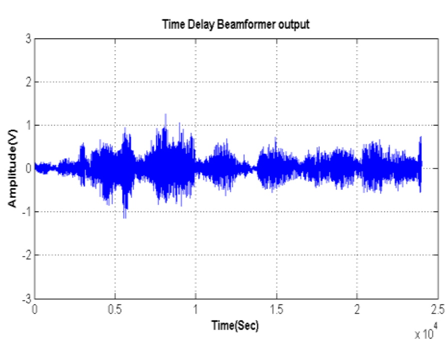

Fig. 4. The Delay and Sum technique output signal.

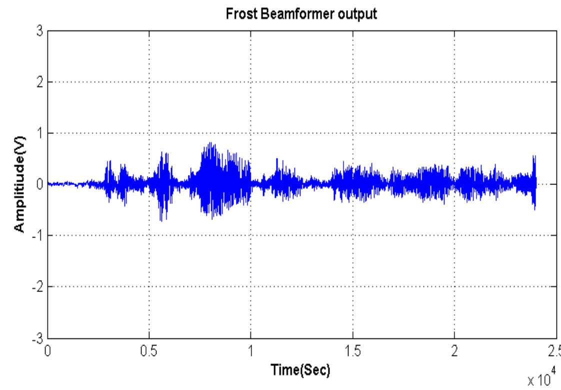

Fig. 5. The Frost Beamformer technique output signal.

Secondly, the influence of the microphone numbers (number of sensors, e.i. the speech enhancement by the array gain) on suppression performance is analyzed.

The plot of output signal-to-interference -plus-noise ratio (SINR) to input SINR under different number of sensors is shown in Fig. 6. The simulation is run as follows: the number of sensors varies from 3 to 23, adaptive frost beamformer each sensor branch had 10 taps, and other parameters are the same as above.

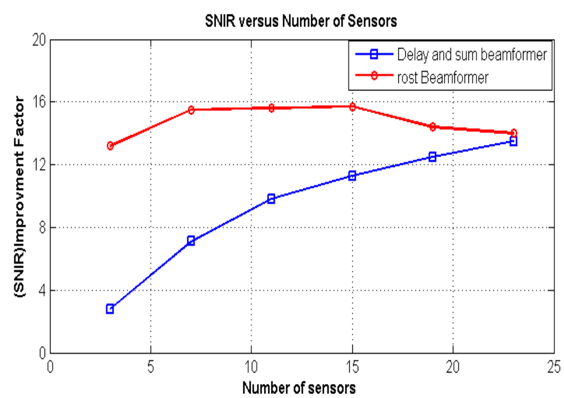

Fig. 6. The SNIR versus different number of sensors.

It can be seen from Fig.6 that, this method can improve Signal to Interference-plus-Noise-Ratio. As per simulation result, as the number of sensors increase from 3 to 23 the SNIR increase rapidly for delay and sum method that is because this method highly depends on the number of sensors. Whereas for the frost beamformer method does not have high influence on sensors, thus there is no remarkable enhancement with increasing the number of sensors. Even though, the frost beamformer had better performance. But, has a high computational cost.

VII. Conclusion

From the above description, we observed that beamforming can enhance signals from the desired direction while preventing ones from other directions. Thus, beamforming can be used for both noise suppression and de-reverberation. Furthermore, it increases the SNR of the output signal which it is the most significant feature to accomplish spatial judgment of speech signals and at the same time defeating uncorrelated noise. It can be observed that, the improvement is more in case of the frost beamformer than the simple Delay and sum beamformer.

REFERENCES

[1] Michael Brandstein and Darren Ward, "Microphone Arrays: Signal Processing Techniques and Applications", Springer, Germany, 2001.

[2] Colin E. Cherry, "Some experiments on the recognition of speech, with one and with two ears", Journal of the Acoustical Society of America, vol. 25, no 5, pp. 975–979, Sep. 1953.

[3] Simon Haykin and Zhe Chen, "The cocktail party problem", Journals of Neural Computation, vol. 17, no. 9, pp. 1875-1902, 2005.

[4] Jacob Benesty, Shoji Makino and Jingdong Chen, "Speech Enhancement", Springer, Germany,2005.

[5] Barry Van Veen and Kevin M. Buckley, "Beamforming: a versatile approach to spatial filtering"IEEE Signal Processing Magazine, vol. 5, pp. 4–24, 1988.

[6] O.L. Frost III, An algorithm for linearly constrained adaptive array processing," Proc. IEEE, vol. 60, pp. 926-935, Aug. 1972.

[7] Vikas.C.Raykar , "A Study of a various Beamforming Techniques And Implementation ofthe Constrained Least Mean Squares (LMS) algorithm for Beamforming " Master thesis 2001.