ISCA Archive
http://www.isca-speech.org/archive

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

PERFORMANCE COMPARISON OF SEVERAL ADAPTIVE SCHEMES FOR MICROPHONE ARRAY BEAMFORMING

Enrique Masgrau, Luis Aguilar and Eduardo Lleida

Dept. of Electronic Engineering and Communications, CPS, University of Zaragoza,
E50015 Zaragoza, SPAIN
E-mail: masgrau@posta.unizar.es, 183427@cepsz.unizar.es, lleida@posta.unizar.es

10.21437/Eurospeech.1999-576

Abstract: The use of microphone array can be a valuable resort for robust speech acquisition. This paper describes a comparative study of the performance achieved by three different schemes of adaptive beamforming with microphone arrays: linear constrained beamforming (Frost [1]), AMNOR system [2] and Affes system [3]. In general, these adaptive schemes are suitable for speech enhancement systems in noisy and reverberant environments, and in the presence of other interfering speech sources (multi-talkers environments). This analysis also includes the classical delay and sum (DS) scheme as a reference for comparative purposes. In order to achieve a homogeneous comparison of the performance, several experimental environments was simulated to evaluate the different schemes, consisting of a desired speech source, an interfering speech or white noise source and the presence of reverberation and omnidirectional noise. This analysis includes simulations of uniform linear arrays, harmonically nested arrays, as well as planar configurations (circular and L-shaped arrays). The Affes system shows be the most suitable system in presence of directional interferences, reverberation and omni-directional noise.

I. INTRODUCTION

Hands-free communications-terminals, audioconference, speech recognition and sound recording are some of the speech related applications which need a robust acquisition system when working in real acoustic sceneries with environment noise and reverberation. In those situations, a robust acquisition system will improve the quality of the speech signal in terms of inteligibility, fidelity or recognition rates.

The microphone array makes use of beamforming techniques to fight against the effects of the acoustics environments. The scope of this comparative analysis comprises four beamforming systems including several acoustics environments and different geometries of the microphone array. First, a time invariant case: the classical delay and sum scheme, included for comparative purposes as a reference. Next, three adaptive beamforming algorithms: linear constrained beamforming (Frost [1]), AMNOR (Adaptive Microphone-array for Noise Reduction) system [2] and Affes system [3]. The features of the different beamformers will be analyzed using two objective measures besides of subjective appreciations: the classical signal-to-noise or interference ratio (SNR) and the signal distortion. This is a resume of a very much extensive work exposed in [4]. In Sections II up V we present some brief reviews of the four algorithms and its more representative performance results. In Section VI the system performances are compared and discussed.

II. DELAY AND SUM BEAMFORMER

From the studied systems, the simplest one is the DS beamformer, which permits to aim the array at only one direction. As it is well known, this system consists in a coherent and uniform weighting sum of the signal component sensors in the desired direction of arrival (DOA). The in-phase signal correction at the desired direction is obtained by a proper delay of the signal sensor. For a fixed number of sensors, the mainlobe width and simmetry of the directivity pattern array depends on the frequency and the pointing DOA (decreasing with the frequency and increasing with broadband to endfire motion). Of course, this frequency dependence can be mitigated using a non-uniform nested array, for instance, the harmonically (logarithmically) nested arrays shown in the figure 1. As it is shown, three uniform subarrays are defined for the respective range band. For narrowband speech (4 KHz) these subarrays and the respective bands are: (b) high band (2-4 KHz), (c) medium band (1-2 KHz) and (d) low band (0-1 KHz). The distance d=4,1cm is chosen to avoid spatial aliasing up to 4000 Hz. The signal in each sensor is filtered to fit the frequency range of each subarray.

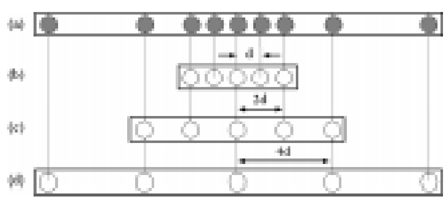

Figure 1. Scheme of a harmonically nested array with a total of 9 sensors; 5 sensors in each subarray

Delay and sum beamforming requires the knowledge of desired DOA. It is the optimal solution in case where the only interfering signal is omnidirectional noise. It does not fight against interferences specifically and, therefore, it is not adequate in presence of directional interferences or reverberation (coherent interferences).

III. LINEAR CONSTRAINED BEAMFORMER

The system with spatial constraints developed by Frost [1] is capable of aiming at the desired directions with a chosen frequency response while canceling the noise contributions, which improves the DS features remarkably. The scheme of the Frost beamformer is shown in figure 2. It consists of M sensors and M FIR filters whose coefficients are constrained to fit the chosen response. This figure shows the equivalent filter response in the look direction as well. The algorithm details can be consulted in [1].

The exact fulfilment of the constraints in the look direction assures a null distortion of the desired signal while maintaining a high cancellation of the directional interferences.

As the DS case, the performance of the Frost beamformer can be greatly improved by using the harmonical array of figure 1. In this case the SNR gains are increased and its values result be very much independent of the factor such input SNR, FIR length and desired DOA.

The Frost system assumes that the signals received at the array come from the direct path between the desired source and the sensors, and hence it is not suitable for multipath or reverberant environments. In this last case, the array will point to as the direct ray as the echo ones, correcting its incoming phases in order to adjust an apparent null at the desired DOA. This fact causes a great attenuation of the desired signal, resulting in a unacceptable signal distortion.

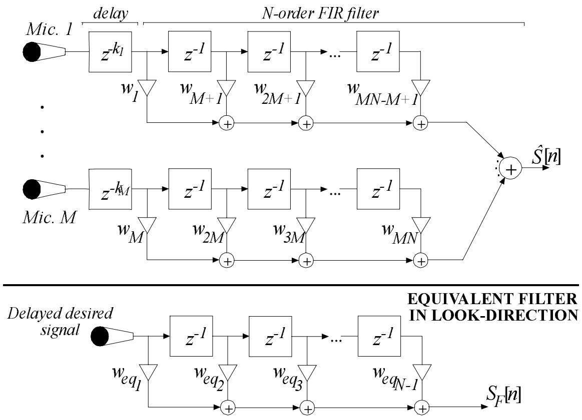

Figure 2. Scheme of the Frost beamformer

To illustrate this cancellation effect we simulated the following scenery: desired speech signal incoming by the broadside (0º); the interference signal consists in the same speech signal incoming by 45º with 5 dB of attenuation (it simulates an extreme reverberation environment consisting in a whole coherent multipath); omni-directional noise at –25dB in relation to the desired signal level; uniform linear array with 5 sensors and FIR length of 8 coefficients. In the figure 3 is shown the directivity diagrams obtained for three representative frequencies: 750, 1750 and 2500 Hz. It can be noted that these diagrams present an unity gains at the 0º and 45º roughly, being in opposition its respective phases. It causes a null in the array output.

The observed peaks presenting gains greater the unity are characteristic of the Frost beamformer when the omni-directional noise level is low (in this case, -25 dB with regard to desired power).

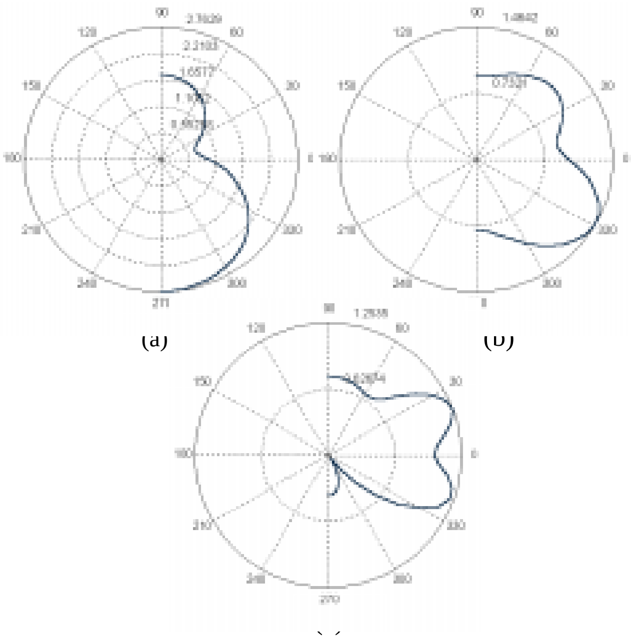

Figure 3. Frost array gains with extreme reverberation to the frequencies a) 750 Hz, b) 1750 Hz, c) 2500 Hz.

IV. AMNOR BEAMFORMING

The AMNOR [2] system introduces a “pilot” desired signal, consisting of white noise, with the same spatial characteristics as the desired speech source. A simplified scheme of this beamformer is shown in the figure 4.

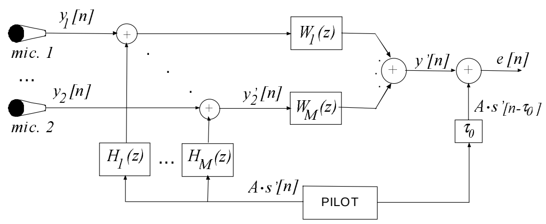

Figure 4. Scheme of the AMNOR beamformer

For this purpose, it is necessary to know exactly the impulse responses from the desired source to each of the microphones. The pilot signal can be a white noise or a tones train with random phases. This pilot is added to the incoming signal and, delayed properly, it is used as reference signal of the beamforming algorithm. The used MMSE criterion leads to a balance between the aiming at the desired DOA, forced by the pilot signal, and the cancelling of the interferences. This balance is drived by the pilot amplitude, in such a way that high amplitude leads to a better aiming of the desired DOA (low distortion) and worse interferences cancelling. The opposed behaviour is presented if the pilot amplitude is low. This behaviour is illustrated clearly in figure 5, where the array frequency response is shown as the desired DOA (0º) and the interference DOA (45º) for two different pilot levels: 5 dB and –15 dB in relation to desired signal power. The experiment scenery is the same as the Frost example but now the interference is uncorrelated with the desired signal

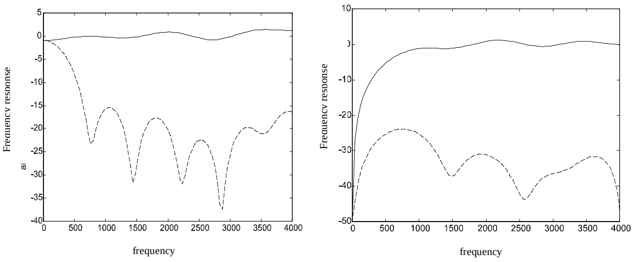

Figure 5. Frequency response of the AMNOR beamforming in the desired DOA (solid line) and interfe-rence DOA (dotted line) for the pilot levels of a) 5dB and b) –15 dB

Thus, this system enables us to set a priori the maximum permissible value of the distortion suffered by the looked source. It exists a tradeoff relationship between the permissible distortion and the SNR gain. The AMNOR algorithm includes the valuable ability of adapting the pilot amplitude to reach the chosen SNR-distortion balance.

In case of knowing the acoustic impulses exactly, this system is suitable for reverberant environments with presence of other interfering signals. This is true due to the signal cancellation effect is reduced because the overall system response to the desired DOA must be close to unity to approach the pilot amplitude, within the refered tradeoff between distortion and interference cancellation. This system overcomes the limitations of the algorithm presented by Frost and achieves better SNR gains, although the impulse responses between the desired source and the arrays should not change; therefore, it is not adequated for variant sceneries. The experiments carried array in this work confirmed its ability to achieve high SNR gains while keeping the distortion below a desired value.

IV. AFFES BEAMFORMING

The Affes system [3] shown in the figure 6 uses a GSLC (Generalized Sidelobe Canceller) scheme in addition of a main beamformer pointing to desired DOA by using a matched filter and sum scheme. Thus, its complexity is fairly high.

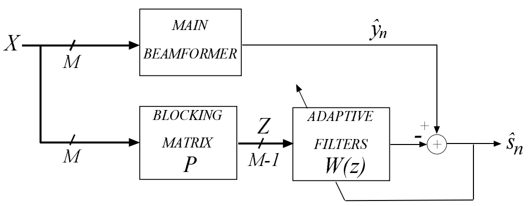

Figure 6. Simplified scheme of Affes system.

This system provides high SNR gains while maintaining a low distortion in the looked speech source. In the figure 7 is shown the directivity diagrams of the main beamformer alone and the overall Affes system for a non-reverberant scenery: desired DOA of 0º and –5dB interference signal with DOA of 45º. Can be see that the main beamformer points to desired DOA but it does not shape nulls in interference DOA. The overall system gets both objectives perfectly. In the figure 9 is shown the frequency response of the overall Affes system to the before example.

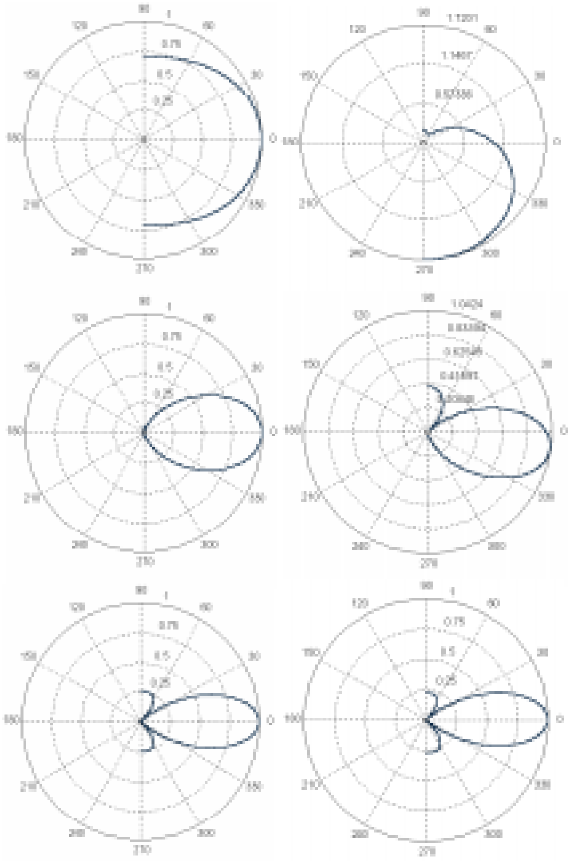

Figure 7. Affes directivity diagrams of the main beamformer alone (left) and the overall system (main beamformer+GSLC) to 750, 1750 and 2500 Hz (top to bottom)

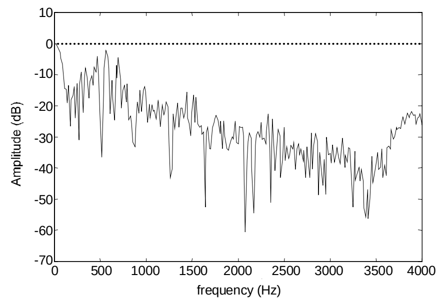

Figure 8.Frequency response of Affes beamformer in the desired DOA (dotted line) and interference DOA (solid line) corresponding to right diagrams of figure 6.9.

As the AMNOR system, the Affes system requires an “a priori” knowledge of the impulse responses between the desired source and the microphones, but this system includes a valuable algorithm to track little movements of the desired source, which showed to work fairly well in our experiments. Thus, this is a very suitable system in lightly variant sceneries with presence of directional interferences, reverberation and omni-directional noise.

To illustrate this ability we simulate a variant reverberant scenery consisting in: interference DOA of 45º, DOA of the desired source of –60º and placed to 1 meter from the array and moving 10 cm around the initial position. The reverberation time of the scenery is 200 ms. roughly. We consider two working cases: tracking ability of the scenery motion OFF and ON. The figure 9 (left) shows the time evolution of the output power of the desired and the interference signals while the GSLC beamforming is adapting with motion tracking ability OFF. A noteworthy cancellation of the desired signal can be observed in addition to a high interference attenuation. On the contrary, in figure 9 (right) corresponding to the case with motion tracking ability ON we can note a almost null signal cancellation while a high interference attenuation is remained. As a resume, the Affes beamformer presents a complexity fairly high, provides high SNR gains while maintaining a low distortion in the looked speech source although working in high reverberant environments. Also, it posses the ability to track little movements of the desired source unlike the rest of the analysed algorithms.

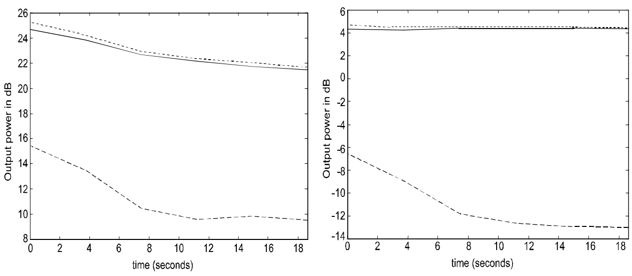

Figure 9. Output power of the desired (solid line), interference (discontinuous line) and global (dots line) with tracking ability OFF (left) and ON (right).

V. CONCLUSIONS

This paper presents a resume of a very much extensive work exposed in [4]. The principal objectives of this work were to carry out a comparative analysis of the principal features of four known beamforming algorithms working in several environments. Firstly, we analysed the time invariant DS system. It requires the knowledge of desired DOA and it is the optimal solution in case where the only interfering signal is omni-directional noise. It is not able to fight against interferences specifically and, therefore, it is not adequate in presence of directional interferences or reverberation (coherent interferences). Lastly, it is necessary to implement the delay compensation blocks, which requires a high sampling frequency or a broadband delay. Then, we present the study of three adaptive beamforming.

The Linear Constrained system due to Frost [1] is capable of pointing at the desired directions with a chosen frequency response while cancelling the interference noise (directional and omni-directional) which improves the DS features largely. It provides high SNR and null distortion of the desired signal at a low complexity cost in non-reverberant environments. This system assumes that the desired signal arrives at the array from a direct path coming on the desired source. Hence, it not suitable for multipath or reverberant environments. In this cases, the beamformer reveals unacceptable signal cancellation effects.

The AMNOR system [2] acquires the desired DOA information by introducing a pilot white signal with the same spatial characteristics as the desired one. In order to work well, the AMNOR system needs to know exactly the acoustic impulse response from the desired source to each of the microphones. On the other hand, this system enables us to set a priori the maximum permissible value of the distortion suffered by the looked source. It exists a tradeoff relationship between the permissible distortion and the SNR gain. The AMNOR algorithm includes the valuable ability of adapting the pilot amplitude to reach the desired SNR-distortion balance. This system overcomes the limitations of the algorithm presented by Frost and achieves better SNR gains, although the impulse responses between the desired source and the arrays should not change; therefore, it is not adequated for variant sceneries. The experiments carried array in this work confirmed its ability to achieve high SNR gains while keeping the distortion below a desired value although in a reverberant environment.

The last studied system is the Affes one [3]. It uses a GSLC scheme, and its complexity is fairly high. This systems provides high SNR gains while maintaining a low distortion in the looked speech source. As the AMNOR system, it requires an a priori knowledge of the impulse responses between the desired source and the array microphones, but this system also includes an algorithm to track little movements of the desired source, which showed to work fairly well in our experiments. Thus, this the most suitable system in presence of directional interferences, reverberation and omni-directional noise. For all systems, it is made simulations with uniform and harmonical linear and planar arrays. The uniform linear arrays behaves worse for low and medium frequency. On the contrary, the harmonical array proves to be very effective to obtain directivity diagrams almost independent of the frequency. Lastly, the planar arrays present a good behaviour and they allow a better discrimination among sources.

[1] Otis L. Frost, “An algorithm for Linearly Constrained Adaptive Array Processing”, Proc. IEEE, vol. 60, n. 8, pp 926-935, August 1972

[2] Y. Kaneda and J. Ohga, “Adaptive microphone array system for noise reduction”, IEEE Trans. on Acoust., Speech, and Signal Processing, vol.34, pp 1391-1400, December1986.

[3] S. Affes and Y. Grenier, “A Signal Subspace Tracking Algorithm for Microphone Array Processing of Speech”, IEEE Trans.on Speech Audio Processing, vol.5, No.5, pp 1063-1071, September 1997.

[4] L. Aguilar “Conformado Adaptativo de Haz en Arrays de Micrófonos para Tratamiento Robusto de Voz”, Career Thesis, CPS, University of Zaragoza, December 1998.