AN ADAPTIVE MICROPHONE ARRAY WITH LOCAL ACOUSTIC SENSITIVITY

Jianfeng Chen, Louis Shue, Hanwu Sun, Koksoon Phua

Institute for Infocomm Research
21 Heng Mui Keng Terrace, Singapore 119613
Email: {jfchen,lshue,hwsun,ksphua}@i2r.a-star.edu.sg

ABSTRACT

In this paper, a microphone array with 3-D focal zone is proposed. The microphone array consists of one omni-directional and two uni-directional microphones. The microphone array is so constructed that a cross zone is formed such that only the sound within this zone is captured and any interferences outside the zone are effectively cancelled. The proposed framework is flexible in defining the location/size of the closed volume where the sound source of interest is located. Simulations have been carried out to demonstrate the 3-D spatial selectivity as well as the noise cancellation performance. The most important feature which differs from the previous works is that the super volumetric selectivity is realized by strategically use only three microphones, by which the overall apparatus acts as a virtual wireless close-talking microphone with confined position constrained in both distance and directions.

1. INTRODUCTION

Effective acquisition of sounds in noisy environments has been a long sought-after goal. In many applications, some of the challenges can be overcome by using close-talking microphones. However, close-talking microphones have some obvious limitations, namely in certain circumstances it is impossible for the user to wear a microphone or for the sound source to be near the microphone. At the same time, audio-capturing at a distance has been tackled using directional microphones to attenuate the interferences. While the directionality in acquiring desired sounds has improved, directional microphones are generally unable to take into account the distance information from microphone to the sound source. In other words, there is no local acoustic selectivity. To date, how to capture the local sound source while excluding other interferences remains an open problem in acoustic engineering as well as communications systems.

Microphone array technologies have been developed for more than two decades [1] and spatial filtering techniques, also known as beamforming, have been intensively studied for target signal enhancement. Among these solutions, the adaptive beamformer [2–4] is regarded as an effective way to achieve higher interference cancellation than the conventional delay-and-sum fixed beamformer. In recent years, adaptive beamforming has also been customized for small microphone array systems, commonly referred as adaptive noise canceling microphones [5–8] whereby only few microphones – as few as two – are used. The performances of these solutions have been studied in [8,9] in terms of their noise cancellation under three typical noise fields, namely coherent, uncorrelated and diffuse noise fields. These solutions are attractive because of their small dimensions, lower computational complexity and reasonable noise reduction performance.

Generally speaking, current research work in this area is still focusing on improving the robustness and the directional selectivity of the system (e.g. [3]) with the problem, no local acoustic selectivity, remains unsolved. In fact, we note that only few works have touched on this issue. In [10], a matched filter array claiming to have the ability of 3-D local acoustic selectivity was presented. The main limitations of this method are that it requires a large number of sensors (ranging from tens to thousands) and the exact locations of the sound source and the microphone array sensors should be known as *a priori*. In [11] Elko proposed a scheme to generate volumetric acoustic focus by using two planar microphone arrays. Based on an estimation of the coherence function, the scheme is somewhat similar to time-varying filtering, which allows the signals correlated to both array outputs to pass and filters out the uncorrelated components. This scheme rests on the assumption that the signal of interest is a stationary ergodic random process and therefore is likely to have difficulty in dealing with speech signal, which is highly non-stationary. Yet in another work, a virtual wireless microphone [12] is proposed to realize the acoustic focus. Delay-and-sum beamformer was employed and the acoustic focus at the desired location was formed by setting gain and delay of each microphone properly. A similar approach has also been attempted in [13] whereby a large aperture array has been used to tackle this problem. By using this method, a good performance is highly dependent on large array dimensions, number of sensors which needs to be large and perfect array calibration.

In this paper, a microphone array based on adaptive noise cancellation is presented. A 3-D local acoustic selectivity is realized by using three microphones, with one omni-directional and two cardioid directional microphones. The microphones are arranged in such a way that the two notches of the cardioid microphones are crossed in a predefined zone. This makes any sound sources within this zone have a relatively weak response in these two microphones, whereas they may still have normal response in the omni-directional one. On the other hand, the sound sources outside the cross zone would have relatively similar response in all the three outputs. By designing an appropriate adaptive noise cancellation system, it is expected that an audio capturing system with 3-D focal zone can be realized.

The paper is organized as follows. In Section 2, the proposed array structure and the adaptive algorithm are described. Some simulation results showing the spatial selectivity and interference cancellation are given next in Section 3. The paper is concluded in Section 4.

2. PROPOSED ARRAY SCHEME

In this section, the proposed framework will be detailed.

0-7803-9332-5/05/$20.00 ©2005 IEEE

2.1. Physical Structure and Formation of Focal Zone

As shown in Figure 1, the microphone array physically consists of three microphones: one omni-directional (M_O) and two uni-directional (M_L, M_R) microphones, which are located at $(0,0)$, (a_L, b_L) and (a_R, b_R) respectively. The directionality of M_O is a circle and that of M_L and M_R is a cardioid with a null in a specific direction. An ideal spatial response $D(\theta)$ for the uni-directional microphone is defined as

$$D(\theta) = \frac{1}{2}(1 - \cos\theta) \tag{1}$$

where θ is the angle between the null direction and a reference direction, usually taken as the direction of the sound source in practice.

We will adopt the following terms for the rest of our discussions. The volume of space contained in the open angle α along the null direction for the uni-directional microphone is termed a *null sector*. The two uni-directional microphones are steered in such a way that the two null sectors intersect to form a closed *focal zone*, indicated as the shaded region A in Fig. 1. We note that although the region A as indicated in Fig. 1 appears as 2-dimensional, it is in fact a closed 3-D volume. Furthermore, depending on different system requirements, the three microphones can be arranged either in one straight line or in a triangle, either symmetric or asymmetric and the angles α can also be variable. All such arrangements will effect the location and the dimension of the focal zone.

2.2. Adaptive Signal Extraction

Let us suppose that the desired sound source $s(t)$ is located in the focal zone A and all the interferences $n_m(t)$, $m = 1, \ldots, M$, are in the other areas, such as B_1, C_1, etc., as shown in Fig. 1. Let us further suppose that these sound sources are uncorrelated to each other. The null directions of the two uni-directional microphones M_L and M_R are steered at β_L and $\pi - \beta_R$ respectively, relative to the horizontal axis X.

Referring to the signal flow shown in the lower part of Fig. 1, the output of M_O, denoted as $x_O(t)$, acting as the primary signal, is then

$$x_0(t) = s(t) + \sum_{m=1}^{M} n_m(t) \tag{2}$$

On the other hand, the outputs of M_L and M_R, $x_L(t)$ and $x_R(t)$, acting as the reference signals, are given by

$$x_L(t) = \epsilon_L s(t - \tau_L) + \sum_{m=1}^{M} d_L^{(m)} n_m(t - \tau_L^{(m)}) \tag{3}$$

$$x_R(t) = \epsilon_R s(t - \tau_R) + \sum_{m=1}^{M} d_R^{(m)} n_m(t - \tau_R^{(m)}) \tag{4}$$

where ϵ_R, ϵ_L, $d_R^{(m)}$ and $d_L^{(m)}$, $m = 1, \ldots, M$, denote the attenuation factors of all sound sources captured by the microphones M_L and M_R, and τ_R, τ_L, $\tau_R^{(m)}$ and $\tau_L^{(m)}$, $m = 1, \ldots, M$, are the time delays of the sound sources reaching M_L and M_R, all taken with respect to M_O. ϵ_R and ϵ_L are two small values indicating the trivial components of the desired sound source received by M_L and M_R.

The mechanism of our proposed scheme is as follows. Since $s(t)$ is located in the focal zone, hence in the null sectors of both M_L and M_R, while the interferences $n_m(t)$, $m = 1, \ldots, M$, are outside the focal zone, the signals $x_L(t)$ and $x_R(t)$ are expected to contain little $s(t)$ and mainly $n_m(t)$. On the other hand, $x_O(t)$ contains a mixture of all the sound sources. Consequently, $n_m(t)$, $m = 1, \ldots, M$, are the common components between primary signal $x_O(t)$ and at least one of the reference signals, $x_L(t)$ and $x_R(t)$. These common components are to be cancelled out through the adjustment of the filters $\mathrm{W}_L(t)$ and $\mathrm{W}_R(t)$. As shown in Fig. 1, the reference signals $x_L(t)$ and $x_R(t)$ will pass through the adaptive filters $\mathrm{W}_L(t)$ and $\mathrm{W}_R(t)$ individually and subsequently subtracted from the primary signal $x_O(t)$ to result in the output $z(t)$. Finally the output $z(t)$ is minimized, which consists of mainly the desired signal $s(t)$.

In a free sound field, the highly correlated components in the primary signal and reference signals will be adaptively diminished to result in only the $s(t)$, the desired signal. However, in practice where certain levels of reverberations exist, some echoes of $s(t)$ fall into the notch sector of M_L or M_R, and will have some contributions in $x_L(t)$ and $x_R(t)$. This may cause target-signal cancellation, similar to the case for conventional adaptive beamformer [2]. Consequently, the adaptive process needs to be controlled delicately to avoid this phenomenon. In light of this, we have followed the scheme proposed in [3], in which the Norm-constrained Least Mean Square (NLMS) adaptation was proposed to overcome this difficulty.

By means of the NLMS adaptive algorithm, the final output $z(t)$ is defined as

$$z(t) = x_0(t) - W_L(t)^T X_L(t) - W_R(t)^T X_R(t) \tag{5}$$

where N is the tap length of the two adaptive filters $\mathrm{W}_L(t)$ and $\mathrm{W}_R(t)$, $()^T$ denotes the transpose operator and

$$W_L(t) = \begin{bmatrix} w_{L,0}(t) & w_{L,1}(t) & \cdots & w_{L,N-1}(t) \end{bmatrix} \tag{6}$$

$$W_R(t) = \begin{bmatrix} w_{R,0}(t) & w_{R,1}(t) & \cdots & w_{R,N-1}(t) \end{bmatrix} \tag{7}$$

$$X_L(t) = \begin{bmatrix} x_L(t) & x_L(t-1) & \cdots & x_L(t-N+1) \end{bmatrix} \tag{8}$$

$$X_R(t) = \begin{bmatrix} x_R(t) & x_R(t-1) & \cdots & x_R(t-N+1) \end{bmatrix}. \tag{9}$$

The updating rule, in the case of $W_L(t)$, consists of the following:

$$\tilde{W}_L = W_L(t) + \mu \frac{z(t) X_L(t)}{||X_L(t)||^2} \tag{10}$$

$$\Omega = ||\tilde{W}_L|| \tag{11}$$

$$W_L(t+1) = \begin{cases} \sqrt{K \tilde{W}_L / \Omega}, & \text{for } \Omega > K, \\ \tilde{W}_L, & \text{otherwise.} \end{cases} \tag{12}$$

where $|| \cdot ||$ is the square norm operator, μ denotes the step size $(0 < \mu < 2)$, $\tilde{W}_L$ the temporal filter vector, Ω and K the total squared-norm of $W_L(t+1)$ and the norm constraint threshold. If $\Omega > K$, $W_L(t+1)$ is restrained by scaling. Similarly, $W_R(t)$ can be updated using the same steps as in (10)-(12).

The purpose of the parameter K is twofold. Firstly, K restricts the over-adjustment of the filters $\mathrm{W}_L(t)$ and $\mathrm{W}_R(t)$, thereby effectively prevents signal cancellation. On the other hand, the span of the null sector is also restricted through the open angle α, which consequently controls the size and the shape of the focal zone A. Although an explicit relationship between K and α is not available, the effectiveness can be seen in the simulations in Section 3.

3. SIMULATIONS AND DISCUSSIONS

In this section, we will discuss some simulation results using the array structure shown in Fig. 1. The three microphones were located at (0.0, 0.0), (-0.20m, 0.10m) and (0.20m, 0.10m) respectively, with the null directions $\beta_L = \beta_R = \pi/4$. The structure was deliberately kept small with a view that the overall system could be readily used for desktop speech capture and recognition applications. The focal zone was located at a distance of about 0.20m to 0.50m away from the centre of the array.

3.1. Spatial Response: Desired Signal Only

In this section, we will study the spatial response of the proposed scheme within an area of 0.80m $\times$ 0.85m, scanned at increments of 0.025m. For this simulation, the sound source consisting of Gaussian white noise was moved to various locations in the simulated region while the energy was maintained constant – to simulate a sound source of constant volume. The outputs of the sensors followed the conventional rules of sound propagation. The remaining parameters were set as follows: the filter length $N = 128$, the step size $\mu = 0.1$ and the sampling rate was 22,050Hz.

For each sound source position, the signals of the three sensors were generated by considering their directionality, relative distances and angles to the sound source. For each location of the sound source, an output $z(t)$ was generated using (10)-(12) for $\mathrm{W}_L(t)$ and similarly for $\mathrm{W}_R(t)$. The output energy level after convergence is shown in Fig. 2 for all positions for the same sound source. To demonstrate the effectiveness of the parameter K, different values were also used in the simulations. As can be seen, the proposed scheme shows a significant spatial sensitivity in the intersection of the two null sectors, rather than a sector without any distance restraint. There is a distinct fall-off in the sensitivity outside the focal zone.

3.2. Spatial Response: Desired Signal with One Interference

To further demonstrate the performance of the proposed system in spatial selectivity, i.e. canceling the interference outside the focal zone, a second scenario involving one target signal and one interference, located at (0m, 0.30m) and (0m, 0.50m) respectively, were simulated. We emphasize that in this case the interference was located directly behind the target signal and the central sensor M_O. The interference was a Gaussian white noise and the desired signal was a short segment of speech. Same parameters as used in Section 3.1 were used with $K = 50$. The output $z(t)$ is shown together with the original speech and the outputs of the three sensors.

It can be seen that the interference was attenuated after about 250ms and the processed signal is much closer to the original and noise-free signal. Similar simulations have been done to test the performance of the proposed method when the target signal and the noise are in other locations. The results are consistent and follow the spatial response as shown in Fig. 2.

Finally, we note that since only three sensors are used, the degree of freedom of the system is relatively small. This in turn suggests that it will be difficult to achieve good performance in a highly reverberant environment and/or one with complicated noise arrangements. Nevertheless, the proposed method provides a relatively simple and efficient solution to realize 3-D focal zone selectivity by using as few as three sensors. The location and the size of the focal zone can be controlled easily by setting the parameter K in (12) and varying the positions of the three sensors.

4. CONCLUSIONS

In this paper a novel microphone array framework using three microphones has been presented. This array structure together with the NLMS scheme, for performing the noise reduction and to avoid the target-signal cancellation, is capable of 3-D volumetric selectivity. The entire apparatus acts as a virtual wireless close-talking microphone. The main application for this structure is in any circumstances where the requirement is to capture sound from a pre-defined zone. As shown in the simulation results, the proposed method has been shown to possess a volumetric selectivity and flexible in defining the location and the size of the focal zone. Some limitations of this idea have also been pointed out which we will report on in future. A real-time experimental setup is currently pursued by the authors.

5. REFERENCES

[1] M. Brandstein and D. Ward, *Microphone arrays: signal processing techniques and applications*. Springer Verlag, Sept. 2001.

[2] L. J. Griffiths and C. W. Jim, “An alternative approach to linearly constrained adaptive beamforming,” *IEEE Trans. Antennas Propagat.*, vol. 30, no. 1, pp. 27–34, 1982.

[3] O. Hoshuyama, A. Sugiyama, and A. Hirano, “A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters,” *IEEE Trans. Signal Processing*, vol. 47, no. 10, pp. 2677–2684, Oct. 1999.

[4] W. H. Neo and B. Farhang-Boroujeny, “Robust microphone arrays using subband adaptive filters,” *IEE Proc. Visual Image Signal Processing*, vol. 149, no. 1, pp. 17–25, 2002.

[5] G. W. Elko and A.-T. N. Pong, “A steerable and variable first-order differential microphone array,” in *Proc. IEEE Int. Conf. on Acoust., Speech, Signal Processing*, vol. 1, Apr. 1997, pp. 223–226.

[6] F.-L. Luo, J. Yang, C. Pavlovic, and A. Nehorai, “Adaptive null-forming scheme in digital hearing aids,” *IEEE Trans. Signal Processing*, vol. 50, no. 7, pp. 1583–1590, July 2002.

[7] K. Phua, J. Chen, L. Shue, and H. Sun, “Development of a compact 2-sensor adaptive directional microphone,” *Signal Processing*, vol. 85, no. 4, pp. 809–820, Apr. 2005.

[8] J. Chen, L. Shue, K. Phua, and H. Sun, “Theoretical comparisons of dual microphone systems,” in *Proc. IEEE Int. Conf. on Acoust., Speech, Signal Processing*, vol. 4, May 2004, pp. 73–76.

[9] J. Chen, K. Phua, L. Shue, and H. Sun, “A robust adaptive cross microphone array,” in *Proc. IEEE Int. Symp. on Circuits Syst.*, May 2005, accepted.

[10] R. J. Renomeron, D. V. Rabinkin, J. C. French, and J. L. Flanagan, “Small-scale matched filter array processing for spatially selective sound capture,” *J. Acoust. Soc. Am.*, vol. 102, no. 5 Pt. 2, p. 3208, Nov. 1997.

[11] G. W. Elko, “Volumetric selective from two beamforming microphone arrays,” *J. Acoust. Soc. Am. Supplement*, vol. 87, no. 1, p. S3, Dec. 1990.

[12] H. Mizoguoguchi, T. Shigehara, M. Yokoyama, and T. Mishima, “Virtual wireless microphone a novel application of real-time visual tracking and sound signal processing,” in *Proc. SICE’98*, July 1998, pp. 999–1004.

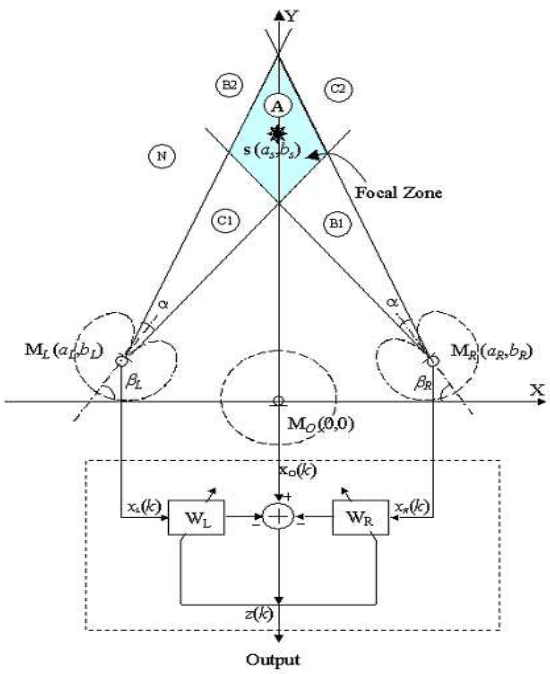

Fig. 1. An illustration of cross notch microphone array system by using two cardioid microphones and one omni-directional microphone as well as adaptive filter structure.

[13] J. M. Sachar, H. F. Silverman, and W. R. Patterson III, "Large vs small aperture microphone arrays: performances over a large focal area," in *Proc. IEEE Int. Conf. on Acoust., Speech, Signal Processing*, vol. 1, May 2001, pp. 3049–3052.

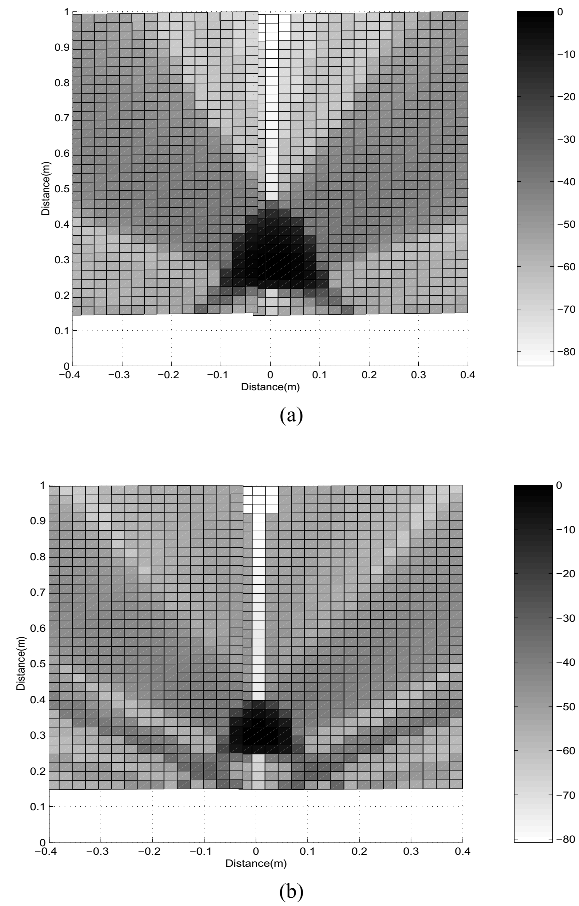

Fig. 2. The spatial responses of the proposed scheme. (a) $K = 50$, (b) $K = 100$.

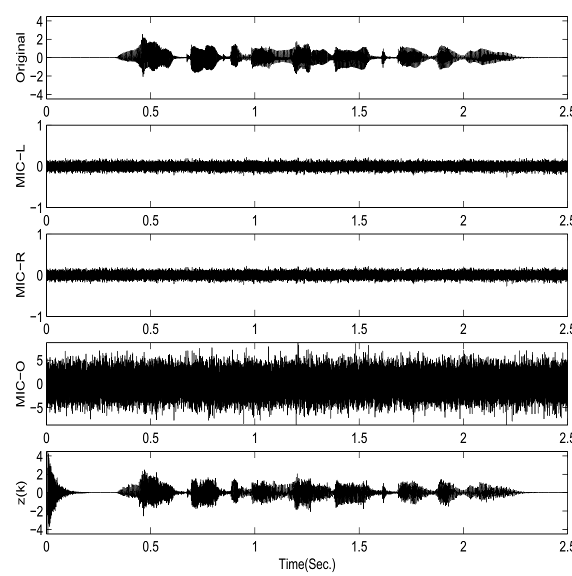

Fig. 3. Original speech, the captured signals at the three microphones and the processed output.