Smart Acoustic Room System and its Application for Car Hand-free Telephone

Rui Chen^{1,a} Fangfang Peng^{1,b} Yuanyuan Luo^{1,c}

¹School of Computer and Information Engineering, Central South University of Forestry & Technology, Changsha, Hunan, China Email: k038656@yahoo.com.cn

Mohammad Reza Alsharif²

²Department of Information Engineering, Faculty of Engineering, University of the Ryukyus, Okinawa, Japan Email: alsharif@ie.u-ryukyu.ac.jp

Abstract - In this paper, a new algorithm of Smart Acoustic Room (SAR) system is prensented for partitioning room acoustically. That is, in different places of the room, one can hear to the desired signal. This is realized by robust estimation algorithm for well control of room acoustic impulse responses. Therefore, unwanted music or speech signal is canceled, while, at the same place, the desired signal (the desired music or speech) could be heard. And also, the application of smart acoustic room system for car hand-free telephone is preseted. By using this new structure, echo and backgroud music in the car will be canceled. In other words, people at far-end will not hear the background sound and echo signal from the near-end room. The computer simulation and experiment results support the theoretical findings and verify the robustness of the proposed algorithms.

Keywords: Acoustic impulse response, Smart Acoustic Room (SAR) system, Hand-free telephone.

I. INTRODUCTION

Study of the room acoustic is an important topic in all kinds of speech processing and audio systems. In hand free telephony or in teleconferencing system, acoustic echo canceller (AEC) [1] is used to remove the echo signal from speech. Here, echo is generated due to acoustic couplage between loudspeaker and microphone in a room. The echo degrades the intelligibility of the communication. Therefore, AEC tries to estimate the room acoustic response and make a replica of the echo signal and remove it. Acoustic noise control (ANC) [2] system is another example to reduce acoustic noise in a location of the room. Here, the acoustic noise is propagated through room acoustic and ANC tries to estimate this acoustic path to generate an opposite signal similar to noise and reduce it appropriately.

In all kinds of above-mentioned examples, we need to estimate and control the room acoustic response between two locations. Nevertheless, this control could be imposed electrically (AEC) or acoustically (ANC), the adaptive digital filter (ADF) is used to perform this job with an appropriate algorithm. Now, we want to introduce a room with smart acoustic (SAR). That is, the acoustic response between two (or more) points could be controlled smartly. By control, we mean to have a well estimation of the acoustic path between two points and then to make the appropriate signal to cancel an unwanted noise or to emphasis to a desired signal (speech or music). In a sense this SAR system works as like as conventional ANC. This is because, we make the control system to be imposed acoustically [3].

In this paper, first the aim of the Smart Acoustic Room system will be introduced. Then we do like to present a simple SAR system by using two speakers and one microphone, the aim of this system is to make a null point at the microphone position locally. That means by control the impulse responses we cannot hear anything at the microphone position. This system includes: the FXLMS [2] based SAR algorithm and SAR algorithm by using virtual microphone [3], [4]. The FXLMS based SAR algorithm is similar to the ANC that estimates the secondary acoustic impulse response (path) by off-line. This solution is not an efficient algorithm. So the Smart Acoustic Room (SAR) system by using the virtual microphone is presented. In this system, the first path and secondary path can be estimated simultaneously. The simple SAR system just can be implemented under a simple single-talk condition, it cannot work under a double-talk condition. As a solution, the SAR algorithm based on correlation function is presented. In this system, there are four speakers output two independent signals. By control the impulse responses smartly, the signals in the room can be separated. We mean to have a well estimation of the acoustic path and then to make the appropriate signal to cancel an unwanted noise or to emphasis to a desired signal (speech or music). Finally we apply the SAR system base on correlation function algorithm [5] into the echo cancellation for car hand-free telephone [6]. That is, by smartly control the impulse response in the room the signals from the loudspeakers will be cancelled at the microphone position. This is a new type echo canceling, it is different with the conventional echo cancellation, which cancel the echo

signals in the telephone system electronically. To demonstrate the validity and the robustness of the proposed algorithms, a lot of simulation and experiments have been done.

II. SMART ACOUSTIC ROOM SYSTEM

A. Aim of Smart Acoustic Room System

As shown as Figure 1, suppose that we want to listen to a Jazz music in one portion of a room and at the same time other fellow wants to listen to a classic one in the other side of the room. Also, we do not want to use headphone as it totally isolate the person from surrounding. Other example is in a conference room or big hall, that we have two kinds of audiences. In one section, audiences want to listen in Japanese while in other section international audiences are seated and they want to listen to the speech in English. Again we do not want to use headphone as here is very costly to manage the system for each person and the Hall should be designed for that or we need transceiver, which is also costly. But if we design the acoustic response such that Japanese loudspeaker covers the desired location while English loudspeaker covers the other part, just by seating in the right place one can hear to desired language. There are much more applications of SAR system. ANC is an especial case of SAR, because in a room we want to reduce the noise source propagation to a location. In more general case, we can define acoustic channels similar as radio or TV channels. Imagine you want to change the channel of TV by using a remote control. The same is possible to be performed for acoustic channel. But, the difference here is location dependency of the remote control. That is, depending on place of the remote control, one can push a bottom to listen to a specific program that be propagated to that place only. If we move the remote control to other location in the room, we can select another program and set the acoustic path to listen only to specified program. Therefore, in SAR we require to change and control acoustic impulse response of the room, as we desire.



Figure.1. Smart Acoustic Room (SAR) system

Of course, sound propagation through acoustic channel from one loudspeaker could cause perturbation for the other one. This is because in contrast to electromagnetic propagation and frequency division multiplexing (by using proper modulation technique) is not possible in acoustic wave propagation. Therefore, by using a powerful algorithm in adaptive digital filter, one can make the null point (zero point) of an acoustic source to be set in specific location and/or move it to any other location. Also, in future work of this class of algorithms the stereo effects [7] will be considered.

B. Definition of SAR System

In Figure 2, a SAR model is shown [3]. The source signal x(n) is for instance a record player output or any audio electric signal. This signal usually converted to acoustic signal through an amplifier and a loudspeaker in order to propagate in a room for listening. The acoustic impulse response from Speaker S_1 to the microphone M is $w_1(n)$ and the one from Speaker S_2 is $w_2(n)$. The aim of this algorithm is to make a null point at the place of microphone M. For this purpose, we put one adaptive filter estimator h(n) in order to predict the acoustic paths and to zero-enforce the signal of M. The signal of microphone is called the error signal, e(n), and it is obtained as follows:

$$e(n) = x(n) * w_1(n) + x(n) * h(n) * w_2(n)$$
(1)
where * denotes the convolution calculation.





Figure .2. SAR model for zero-enforcing microphone. Signal

If h(n) is adapted perfectly, then the error signal will be diminished to zero. That is we can make a null point at the microphone position. Therefore, in Z transform we have:

 $X(z) * W_1(z) + X(z) * H(z) * W_2(z) = 0$

That is:

$$H(z) = -\frac{W_1(z)}{W_2(z)} \tag{3}$$

(2)

 $W_2(2)$ (3) As it shown in Figure 2, similar as the ANC, assuming that the acoustic pass $W_2(n)$ is the secondary path, off-line modeling can be used to estimate $W_2(n)$ during an initial training stage. When the $W_2(n)$ was known, the input of the adaptive filter can be defined as:

$$z(n) = w_2(n) * x(n)$$
 (4)

Then the coefficients of the adaptive filter h(n) can be update by using the LMS algorithm, this solution is called Filtered_X LMS FXLMS algorithm.

$$h_{i}(n+1) = h_{i}(n) - 2\mu e(n)z(n-i)$$
⁽⁵⁾

As we know, the FXLMS algorithm is not an efficient algorithm, because we must estimate the secondary path by off-line. If the secondary path was changed, the error signal will be mislead estimation of the adaptive filter. To solve this problem, the virtual microphone algorithm was presented.

C. Smart Acoustic Room (SAR) System by using Virtual Microphone

In this paper, we challenge to control the acoustic response between two points as shown in Figure.3. That is by using two speakers and one microphone to make an acoustic null point at the microphone position. In Figure.4, a SAR model by using the virtual microphone [4], [5] is shown. The source signal x(n) is for instance a record player output or any audio electric signal. This signal usually converted to acoustic signal through an amplifier and a loudspeaker in order to propagate in a room for listening. The acoustic paths from Speaker S1 to the microphone M is w1(n) and the one from Speaker S2 is w2(n). We want to make a null point at the place of microphone M. For this purpose, first we put one adaptive filter estimator h(n) in order to predict the acoustic paths and to zero-enforce the signal of M. The signal of microphone is called the error signal, e(n), and it is obtained as follows:

$$e(n) = x(n) * w_1(n) + x(n) * h(n) * w_2(n)$$
(6)

Aside of speakers S1 and S2, we imagine that we have two virtual speakers $\tilde{S}1$ and $\tilde{S}2$ in parallel with S1 and S2, respectively. Also, we define two virtual acoustic paths for $\tilde{S}1$ and $\tilde{S}2$ as $\tilde{w1}(n)$ and $\tilde{w2}(n)$ from each virtual speaker to a virtual microphone \tilde{M} (see Fig.2). The signal of the virtual microphone is $\tilde{e}(n)$. According to Fig.2, we can write the following relation for the virtual paths:

$$\widetilde{e}(n) = x(n)^* \widetilde{w}_1(n) + x(n)^* h(n)^* \widetilde{w}_2(n)$$
(7)

If h(n) is adapted perfectly, then the virtual error signal will be diminished to zero. Therefore, in Z transform we have:

$$X(z) * \widetilde{W}_{1}(z) + X(z) * H(z) * \widetilde{W}_{2}(z) = 0$$
(8)

That is:

$$H(z) = -\frac{W_1(z)}{\widetilde{W}_2(z)} \tag{9}$$

From Eq. (1) and (3), we conclude that:

$$\frac{W_1(z)}{W_2(z)} = \frac{\widetilde{W_1}(z)}{\widetilde{W_2}(z)} \Longrightarrow \frac{W_1(z)}{\widetilde{W_1}(z)} = \frac{W_2(z)}{\widetilde{W_2}(z)} = a(z)$$
(10)

Function a(z) describes the relation between the real and virtual part of the system. Then we can use two simple LMS adaptive filters to estimate the impulse responses w1 and w2. For estimation the w1, the error signal can be written:

$$E_{w1}(z) = [W_1(z) - a(z)W_1(z)]X(z)$$

$$= W_1(z)X(z) - \frac{W_2(z)}{\widetilde{W}_2(z)}\widetilde{W}_1(z)X(z) = E(z)$$
(11)



Figure.3. Two speakers SAR system

As the same for estimation the w2., the error signal can be written:

$$E_{w2}(z) = -E(z) \tag{12}$$

That is, the acoustic paths w1(n) and w2(n) can be estimated by using the real error e(n). In order to reduce the computational complexity at this time all the computation will be done in the frequency domain [6]. First, the FFT of the input signals x(n), y(n) are calculated.

$$F_{x}(n, p) = \sum_{k=0}^{N-1} x(n-k) W^{kp}$$
(13)

$$F_{y}(n, p) = \sum_{k=0}^{N-1} y(n-k) W^{kp}$$
(14)

where W shows complex exponential $e^{-j(2\pi/N)}$, N is the impulse response length. Then, the FFT transform of the error signal is calculated.

$$F_{e}(n, p) = \sum_{k=0}^{N-1} e(n-k) W^{kp}$$
(15)



Fig.ure.4. SAR model by using the virtual microphone

So, the acoustic impulse response can be estimated by:

$$\widetilde{W}_{1}(n+1,p) = \widetilde{W}_{1}(n,p) + \frac{2\mu F_{e}(n,p)F_{x}^{*}(n,p)}{1+tr[F_{x}(n,p)F_{x}(n,p)]}$$
(16)

$$\widetilde{W}_{2}(n+1,p) = \widetilde{W}_{2}(n,p) - \frac{2\mu F_{e}(n,p)F_{y}^{*}(n,p)}{1 + tr[F_{y}(n,p)F_{y}(n,p)]}$$
(17)

The superscript * shows the Hermitian transposition and tr[·] means the trace operator. Finally H(z) is calculated by Eq (14) and h(n) can be calculated by using the inverse FFT transform.

D. Smart Acoustic Room System based on Correlation Function

In this section, the SAR algorithm based on the correlation function [8] [9] [10] [11] is presented . The SAR system is shown in Figure.5. The aim of this system is that by control the acoustic impulse responses the signals in the room can be separated. The person can choose the desired signal just by seating at the different position. As shown as in Fig.3, the person who seating at the position A just can hear the desired signals from speakers S3 and S4, because the signals from speakers S1 and S2 was cancelled. The same process will be done for the position B. The person who seating at the position B, just can hear the desired signal from the speakers S1 and S2.

Because the processes for position A and B are same, at here just the process for the position A will be introduced. In the Figure.6, the structure of the proposed SAR algorithm is shown. The desired signal from speakers S3 and S4, are assumed as the double-talk signals. Also the proposed algorithm will be implemented in the frequency domain.

For the double-talk condition the signal from the microphone will be defined as follows:

$$d(n) = e(n) + s(n)$$

$$= x(n) * w_1(n) + x(n) * h(n) * w_2(n) + s(n)$$
(18)

First the auto-correlation of the input signal is calculated:

$$R_{xx}(n,k) = \sum_{j=0}^{n} x(j)x(j-k)$$
 (19)

$$R_{yy}(n,k) = \sum_{j=0}^{n} y(j) y(j-k)$$
 (20)



Fig.ure.5. Four speakers SAR system

And then the cross-correlation function is calculated:

$$R_{dx}(n,k) = \sum_{j=0}^{n} d(j)x(j-k)$$
 (21)

$$R_{dy}(n,k) = \sum_{j=0}^{n} d(j) y(j-k)$$
(22)

The fast Fourier transform is shown as below:

$$F_{xx}(n,p) = \sum_{k=0}^{N-1} \left[\sum_{j=0}^{n} x(j)x(j-k) \right] W^{kp}$$
(23)

$$F_{yy}(n,p) = \sum_{k=0}^{N-1} \left[\sum_{j=0}^{n} y(j) y(j-k) \right] W^{kp}$$
(24)

$$F_{dx}(n,p) = \sum_{k=0}^{N-1} \left[\sum_{j=0}^{n} d(j) x(j-k) \right] W^{kp}$$
(25)

$$F_{dy}(n,p) = \sum_{k=0}^{N-1} \left[\sum_{j=0}^{n} d(j) y(j-k) \right] W^{kp}$$
(26)

So the acoustic paths can be updated by:

$$\widetilde{W}_{1}(n+1,p) = \widetilde{W}_{1}(n) + \frac{2\mu F_{dx}(n,p)F_{xx}^{*}(n,p)}{1+tr[F_{xx}(n,p)F_{xx}(n,p)]}$$
(27)

$$\widetilde{W}_{2}(n+1,p) = \widetilde{W}_{2}(n,p) - \frac{2\mu F_{dy}(n,p)F_{yy}^{*}(n,p)}{1 + tr[F_{yy}(n,p)F_{yy}(n,p)]}$$
(28)

The superscript * shows the Hermitian transposition and $tr[\cdot]$ means the trace operator.



Figure.6. SAR system based on correlation function

E. Application of Smart Acoustic Room System

In this section we applied the SAR system into the car hand-free telephone. In order to verify the robustness of the proposed algorithm, some computer simulations were done.

In this simulation, the car hand-free telephone was used to collect the signal from the speakers S1 and S2 As shown as in Figure.7, we assume that the person was talking by using the hand-free telephone under a music environment in the car. In this condition, the music signal and the echo signal from the far-end room are unwanted. If we had a good estimation of the acoustic paths from the speakers to the microphone, and then could to make the appropriate signal to cancel the unwanted background music signal. Therefore, the people at far-end will not hear the background music and echo signals from the near-end room, the quality of communication is improved.

In this system, two adaptive filters by using the SAR system are used, one is used for the echo canceling, one is used for the background music canceling. The structure of the echo canceller is shown as the Figure.8. The input signal x(n) is the speech signal from the far-end. In this adaptive filter system, the background music signal and the voice of the person who is talking in the near-end room are assumed as the double-talk signal s(n). And also the same structure is used for the background music canceling. The input signal x(n) is the background music signal, the speech signals from the far-end and near-end room are assumed as the double-talk signal s(n). So if the impulse responses w1 and w2 were estimated, the echo and the background signals could be canceled at the microphone position just using one microphone.



Mic is the microphone of the car hand-free telephone





Figure.8. . Echo and Background Music canceling by using SAR system

III. EXPERIMENTAL AND SIMULATION RESULTS

First to demonstrate the validity and the robustness of the SAR algorithm by using the virtual microphone in the real world, some experiments were performed. At first, the real world acoustic impulse responses of 81 positions that we named 1-1 to 9-9 were measured. The distance between two positions is 20cm. The experiment environment is shown as Figure.9. The real world impulse response is 256 taps. Then by using the measured impulse response, some simulations were done. The input signal x(n) is a music sound. The aim is that by using the virtual microphone SAR system to make a null point at position 5-5. It is mean that the acoustic response between two points could be controlled smartly. At position 5-5 we cannot hear anything from the two speakers. And also, the characteristics of the other position around of the position 5-5 were examined.

In the table 1, the Mean Squared Error (MSE) of SAR algorithm by using the virtual microphone is shown. At the position 5-5 the SAR algorithm by using the virtual microphone converges to -35db. That is, by smartly control the acoustic impulse responses, a null point at the microphone position. at this position 5-5 the person almost can not hear anything. And because the similarity of impulse responses, SAR system can make a null point almost 40-60 cm ranges . In the Figure 10, the 3D hole of the MSE is shown.

Then in order to test the algorithm activity, some computer simulation were done. The acoustic paths w1 (n) and w2 (n) of the room were assumed to have exponential decaying shape that decreases to -60dB after M sample, which were defined as follows:

$$w_{1,2}(i) = Randn[exp(-8i/M)]$$
 (29)

where Randn is a normal distributed random number between +1,-1 with zero mean and unit variance. W1,2 are the impulse responses from the speakers S1 and S2 to microphone, respectively.

To measure the performance of the algorithm, also the MSE (Mean Squared Error) is used. The Mean Squared Error (MSE) of the algorithms is shown. The MSE can be defined as:

$$MSE = \frac{1}{M} \sum_{k=1}^{M} e(k)^{2}$$
(30)

The input signal x(n) for the two adaptive filters were shown in Figure.11-a and 11-b, respectively. The input signal of the background music canceling is a music signal as shown in Figure.11-a. For the echo canceling, the input signal is a speech of woman in English as shown in Figure.11-b. In Figure.11-c, the waveform of the error signal at the microphone position is shown. The unwanted signals were canceled at the microphones position locally. Therefore, the people at far-end will not hear the echo and background music signals from the near-end, the quality of communication is improved. In Figure.11-d, MSE of the SAR system was shown. The MSE of the proposed algorithm converged to -38dB.

RoomSize 5.8m \times 3.6m \times 3.0m

Figure.9. Experiment Environment

TABLE 6.1 MSE OF SAR ALGORITHM (DB)

MISE OF SAR ALGORITHM (DD)									
	1	2	3	4	5	6	7	8	9
1	-15	-17-	-19	-21	-22	-18	-18	-13	-15
2	-15	-14	-18	-19	-22	-19	-20	-12	-17
3	-17	-15	-17	-23	-22	-22	-18	-17	-17
4	-16	-22	-19	-20	-18	-21	-23	-21	-21
5	-20	-19	-22	-23	-35	-21	-20	-19	-23
6	-18	-19	-19	-22	-17	-19	-22	-23	-23
7	-22	-19	-19	-17	-21	-21	-19	-22	-22
8	-23	-22	-23	-22	-24	-20	-20	-18	-23
9	-23	-19	-19	-22	-24	-23	-22	-21	-21



Figure.10. 3D hole of the MSE



Figure.11. Waveform of signals

IV. CONCLUSIONS

In this paper, a new class of smart acoustic room sysytem algorithm and its application are presented. By control the impulse responses smartly, we mean to have a good estimation of the acoustic paths from the speakers to the microphone, and then to make the appropriate signal to cancel the unwanted signal. The computer simulation and experiment results support the theoretical findings and verify the robustness of the proposed algorithms

ACKNOWLEDGMENTS:

This project supported by Hunan Provincial Natural Science Foundation of China (No.10JJ5062)"

REFERENCES

- [1]. S. Haykin, "Adaptive Filter Theory, "Third Edition Prentice Hall, 1996.
- [2]. Sen M. Kuo, Dennis R. Morgan, "Active Noise Control Systems", John Wiley & Sons, Inc, 1996.
- [3]. M. R. Alsharif, R, Higa, R. Chen, "Smart Acoustic Room (SAR)", The 2004 Autumn Meeting of the Acoustic Society of Japan, pp. 601-602, Sep.2004.
- [4]. M. R. Alsharif, R. Chen, R. Higa, "Smart Acoustic Room (SAR) System by Using Virtual Microphone", International Symposium on Telecommunications (IST), Shiraz, Iran, pp.683-686, Sep.2005.
- [5]. R., Chen, M. R. Alsharif,, K. Yamashita, "A New Type Echo Canceling by Using the Smart Acoustic Room (SAR) system & Correlation Function For the Double-Talk Condition", IEEE, EURASIP, 9th International Workshop on Acoustic Echo and Noise Control (IWAENC 2005), Eindhoven, The Netherlands, pp.29-32, Sep.2005.
- [6]. Di Wu, Chenxi Hou, Limin Sun, Yan Ling, Jiangchuan Liu, "The XM Satellite Radio Software Module of an Embedded Car Audio System", Journal of Software, Vol 7, No 9 (2012), 1981-1992, Sep 2012

- [7] Li Li, CaiMing Zhang, Hua Yan, "Stereo Matching Algorithm Based on a Generalized Bilateral Filter Model", Journal of Software, Vol 6, No 10 (2011), 1906-1913, Oct 2011
- [8]. Y. Ohana, T. Kohna, "Direct Fully Adaptive Active Noise Control Algorithms Without Identification of Secondary Path Dynamics", IEEE International Conference on Control Application, Scotland, 2002.
- [9]. M. R. Alsharif, T. Hayashi, K. Yamashita, "Correlation LMS Algorithm and its Application to Double-talk Echo Canceling," ELECTRONICS LETTERS, vol.35, No.3, pp.194-195, Feb.1999.



Rui Chen (correspondent author): associate professor, He received the M.Sc. and Ph.D.degree ifrom the University of Ryukyu, Japan, 2003, 2006, respectively. Now he works at School of Computer and Information Engineering, Central South University of Forestry & Technology, china. His research interests are in the fields of echo cancellation, active noise

control, speech synthesis, image processing, wavelet transform, principa component analysis, independent component analysis, watermarking, and image edge detection.



Fangfang Peng: She received the B.E. from Central South University of Forestry & Technology, china, 2009. Now She worked as a master student in the same university. Her research interests are independent component analysis, active noise control.



Yuanyuan Luo: She was born in Huna, China, on July 9,1986. She received the B.E. in computer science from the Central South University of Forestry and Technology, now she is a graduate student in the same university. Her research interests are in the field of digital image processing, automatic recognition and alarm system of forest fire image.

- [10]. M. R. Alsharif, A. Shimabukuro, T.Hayashi, K. Yamashita, "Expanded CLMS Algorithm for Double-talk Echo Cancelling," Proc IEEE, SMC'99, Japan, vol.1, pp. 998-1002, Oct.1999.
- [11]. M. R. Alsharif, R. Chen, K. Yamashita, "A New Class of Adaptive Algorithm Based on Correlation Functions for Double-Talk Acoustic Echo Cancelling," Journal of Mathematics and Computer in Simulation, Elsevier Publisher, vol. 65/6 pp. 599-605,



Mohammad Reza Asharif : was born in Tehran, Iran, on December 15, 1951. He received the B.Sc. and M.Sc. degree in electrical engineering from the university of Tehran, Tehran, 1973, 1974, respectively and the Ph.D. degree in electrical engineering from the university of Tokyo, Tokyo in 1981. He was Head of Technical Department of T.V. broadcasting college (IRIB), Tehran, Iran from 1981 to 1985.

Then, he was a senior researcher at Fujitsu Labs. Co. Kawasaki, Japan from 1985 to 1992. Then, he was an assistant professor in Department of Electrical and Computer Engineering, University of Tehran, Tehran, Iran from 1992 to 1997. Dr. Asharif is now a full professor at Department of Information Engineering, University of the Ryukyus, Okinawa, Japan since 1997. His research interests are in the field of digital signal processing, acoustic echo cancelling, active noise control, adaptive digital filtering, image and speech processing. Professor Asharif is a senior member of IEEE since 1998.}