Proceedings of the Institute of Acoustics

WAVEFRONT RECONSTRUCTION USING OPTIMAL LINEAR FILTERING

P.A. Nelson (1), O. Kirkeby (1), and H. Hamada (2)

(1) Institute of Sound and Vibration Research, Southampton University, England
(2) Acoustic Laboratory, Department of Electrical Communication Engineering, Tokyo Denki University, Japan

1. INTRODUCTION

Here we present an approach to the reproduction of direction-of-arrival information measured by a compact array of sensors used to record acoustic signals. An array of microphones is used to record signals produced by an incident sound field. The objective is to reproduce, as closely as possible, the time histories of the recorded signals together with the direction of propagation of the sound waves producing the signals. It is shown below that a very straightforward technique can be used to accomplish this. The approach taken relies on the construction of a matrix of linear filters which are derived by using a least squares technique. The matrix operates on the vector of recorded signals and produces a vector of signals input to an array of acoustic sources used for reproducing the field. The general scheme is illustrated in block diagram form in Figure 1. The design of the filter matrix $\mathbf{H}(z)$ is accomplished by minimising the mean square error between the desired signals $\mathbf{d}(n)$ and the reproduced signals $\hat{\mathbf{d}}(n)$. The desired signals are simply delayed versions of the original recording. It is shown below that this simple design philosophy leads to source input signals which then result in a good approximation to the directional properties of the recorded sound field being reproduced in a restricted region of space.

2. THE DESIGN OF THE OPTIMAL FILTER MATRIX

The approach taken to the filter design is that specified in reference [1] and which has been used previously in connection with problems in the active control of sound. First the "filtered reference signals" are defined. These are the signals generated by passing the k'th recorded singal $u_k(n)$ through the transfer function $C_{lm}(z)$ which comprises the l,m'th element of the matrix $\mathbf{C}(z)$. This signal is denoted $r_{lmk}(n)$. The generation of the filtered reference signal can be explained with reference to the block diagram of Figure 2. Since the system is linear, the operation of the elements of the transfer functions $\mathbf{H}(z)$ and $\mathbf{C}(z)$ can be reversed. In discrete time, the sampled signal reproduced at the l'th location in the sound field can be written as

$$\hat{d}_l(n) = \sum_{k=1}^{K} \sum_{m=1}^{M} s_{lmk}(n), \tag{1}$$

where the signal $s_{lmk}(n)$ is defined by

WAVEFRONT RECONSTRUCTION

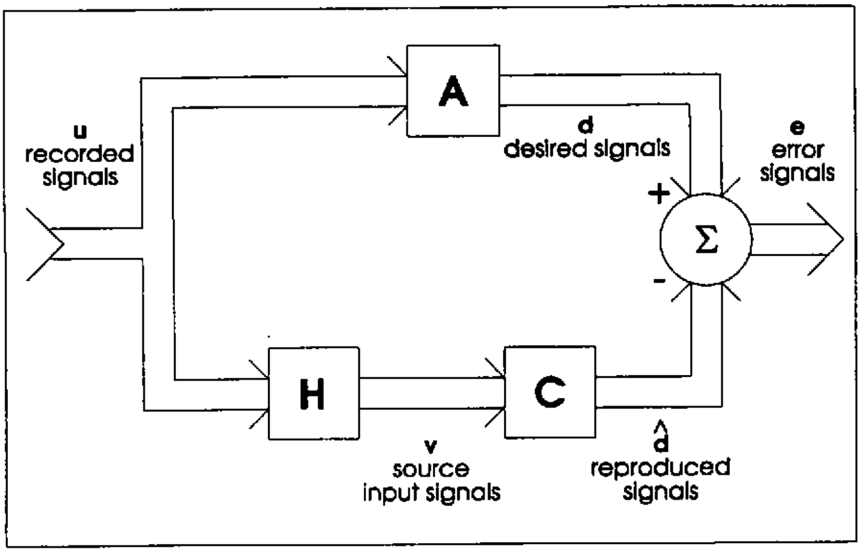

*Figure 1. The sound reproduction problem in block diagram form. The vector **u** is a vector of recorded signals, **v** is a vector of signals input to the sources used for reproduction and $\hat{\mathbf{d}}$ is a vector of signals reproduced in the sound field. The vector **d** defines the vector of signals that are desired to be reproduced and **e** = **d** - $\hat{\mathbf{d}}$ is a vector of error signals. The matrix **C** defines the transfer functions between **v** and $\hat{\mathbf{d}}$, and the matrix **H** defines a matrix of filters which are used to operate on the recorded signals **u** in order to determine the source input signals **v**. The matrix **A** is used to define the desired signals **d** in terms of the recorded signals* **u**

$$s_{lmk}(n) = \sum_{i=0}^{I-1} h_{mk}(i) r_{lmk}(n-i), \tag{2}$$

and $h_{mk}(i)$ is the i'th coefficient of the FIR filter processing the k'th recorded signal to produce the m'th source input signal. Each of the FIR filters is assumed to have an impulse response of I samples in duration. Thus the signal $\hat{d}_l(n)$ can also be written as

$$\hat{d}_l(n) = \sum_{k=1}^{K} \sum_{m=1}^{M} \mathbf{h}_{mk}^{\mathrm{T}} \mathbf{r}_{lmk}(n), \tag{3}$$

where the vectors h_{mk} and $r_{lmk}(n)$ are defined by

$$\mathbf{h}_{mk}^{\mathrm{T}} = \left[h_{mk}(0) \quad h_{mk}(1) \quad h_{mk}(2) \ldots h_{mk}(I-1) \right], \tag{4}$$

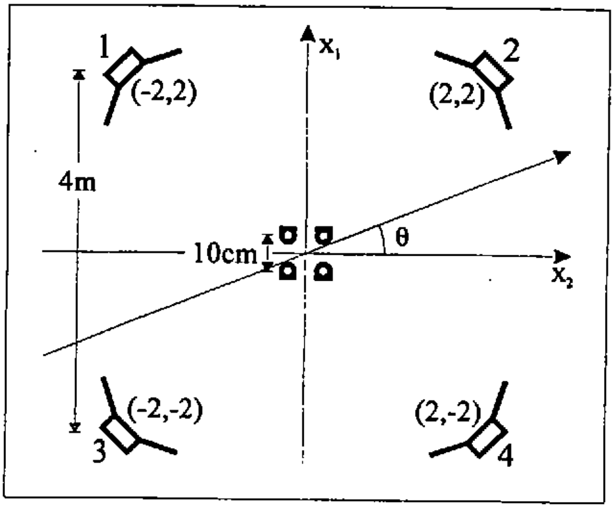

Figure 2. The geometrical arrangement of reproducing sources and recording transducers used for the design of a causal, stable realisation of the optimal filter matrix H. *Four sources were used in the coordinate positions shown together with four sensors spaced 0.1 m apart on a square grid*

$$\mathbf{r}_{lmk}{}^{T}(n) = \left[r_{lmk}(n) \quad r_{lmk}(n-1) \quad r_{lmk}(n-2) \ldots r_{lmk}(n-I+1) \right]. \tag{5}$$

The following composite vectors are now defined

$$\mathbf{h}^{T} = \left[\mathbf{h}_{11}{}^{T}\, \mathbf{h}_{12}{}^{T} \,..\, \mathbf{h}_{1K}{}^{T} | \mathbf{h}_{21}{}^{T}\, \mathbf{h}_{22}{}^{T} \,..\, \mathbf{h}_{2K}{}^{T} | \ldots | \mathbf{h}_{M1}{}^{T}\, \mathbf{h}_{M2}{}^{T} \,..\, \mathbf{h}_{MK}{}^{T} \right], \tag{6}$$

$$\mathbf{r}_{l}{}^{T}(n) = \left[\mathbf{r}_{l11}{}^{T}(n) \,..\, \mathbf{r}_{l1K}{}^{T}(n) | \mathbf{r}_{l21}{}^{T}(n) \,..\, \mathbf{r}_{l2K}{}^{T}(n) | \ldots | \mathbf{r}_{lM1}{}^{T}(n) \,..\, \mathbf{r}_{lMK}{}^{T}(n) \right], \tag{7}$$

$$\hat{\mathbf{d}}^{T}(n) = \left[\hat{d}_1(n) \quad \hat{d}_2(n) \quad \hat{d}_3(n) \ldots \hat{d}_L(n) \right], \tag{8}$$

together with the matrix

$$\mathbf{R}^{T}(n) = \left[\mathbf{r}_1(n) \quad \mathbf{r}_2(n) \ldots \mathbf{r}_L(n) \right]. \tag{9}$$

These definitions are used in reference [2] to find the solution for the optimal set of coefficients in the composite vector **h** that minimises the time averaged sum of squared errors between the desired and reproduced signals. The cost function minimised is given by

WAVEFRONT RECONSTRUCTION

$$J = E\left[\mathbf{e}^T(n)\mathbf{e}(n) + \beta\,\mathbf{v}^T(n)\mathbf{v}(n)\right], \qquad (10)$$

where the error vector $\mathbf{e}(n) = \mathbf{d}(n) - \hat{\mathbf{d}}(n)$ and the second term in the cost function weights the effort associated with the source input signals $\mathbf{v}(n)$. If all the recorded signals comprising the vector $\mathbf{u}(n)$ are assumed to be mutually uncorrelated white noise sequences with a mean square value of σ^2, it can be shown that the optimal composite tap weight vector that minimises J is given by

$$\mathbf{h}_0 = \left\{E\left[\mathbf{R}^T(n)\mathbf{R}(n)\right] + \beta\sigma^2\mathbf{I}\right\}^{-1} E\left[\mathbf{R}^T(n)\mathbf{d}(n)\right]. \qquad (11)$$

Equation (11) therefore defines the optimal values of all the coefficients in the filters that comprise the matrix **H**. One way to determine these coefficients is obviously by direct inversion of the matrix in equation (11). However, this matrix is clearly of high order, being of dimension $I \times M \times K$. Another approach is to use the LMS algorithm, extended for use with multiple errors by Elliott and Nelson [3,4]. It can be demonstrated that the algorithm can be written in the form

$$\mathbf{h}(n+1) = \gamma\,\mathbf{h}(n) + \alpha\,\mathbf{R}^T(n)\mathbf{e}(n), \qquad (12)$$

where α is a convergence coefficient and γ is a "leak coefficient" whose value is directly related to the penalisation of effort associated with the parameter β.

3. RESULTS OF COMPUTER SIMULATION

Some results of simulation which use this filter design technique will be presented here. This technique has been used to design a causal, stable realisation of the filter matrix $\mathbf{H}(z)$ used to operate on the signals recorded by four sensors in order to provide the inputs to four sources used to "reconstruct optimally" the direction of the arrival of the waves in the region in which the recordings were made. Note that the four sensors are placed in a square array of dimension 0.1 m, as illustrated in Figure 2. The effective sample rate used was 34kHz. This enabled the matrix $\mathbf{C}(z)$ to be approximated to good accuracy by transfer functions of the form of $C_{lm}(z)=\rho_0 z^{-\Delta lm}/4\pi R_{lm}$ with Δ_{lm} given by the closest integer value to R_{lm}/c_0, where ρ_0 and c_0 are the density and sound speed. The delays Δ_{lm} were all in the range between 270 and 290 samples and the matrix $\mathbf{A}(z)$ was assumed to be $\mathbf{I}z^{-\Delta lm}$ with the modelling delay delta set equal to 350 samples. Each of the filters in $\mathbf{H}(z)$ was assumed to have 128 coefficients. Having designed these filters by using the algorithm in equation (14), their effectiveness in producing the appropriate value of $v_l(n)$ was evaluated by assuming that the recorded signals $u_1(n)$ to u_4(n) were produced by plane waves falling on the sensor array at an angle θ (Figure 2). The waves were assumed to produce a white noise sequence, with a power spectral density of unity, the same sequence being recorded by each sensor but all differing by delays that are a function of only θ. The power spectral density of a given source input signal could then be calculated. Figure 3 shows the power spectral density of the input signal to source 1 as a function of both frequency and the angle of incidence θ of the recorded waves. Clearly at very low frequencies (45Hz), the source produces an output irrespective of θ, which one might anticipate when the distance between the sensors is very small compared to the wavelength of the incident field. At frequencies up to about 1500Hz, the source only produces an output for waves falling in the range of angles of incidence which can effectively be reproduced by the source. Above this frequency, the effect of inadequate spatial sampling of the field becomes apparent and the source will produce an output for waves having angles of incidence that the source cannot hope to reproduce. These results emphasise the requirement to comply with the sampling theorem by having the recording sensors spaced apart by less than one half (and preferably one third) of an acoustic wavelength at the highest frequency of interest. Nevertheless, the results show considerable promise and the technique clearly offers scope for refinement.

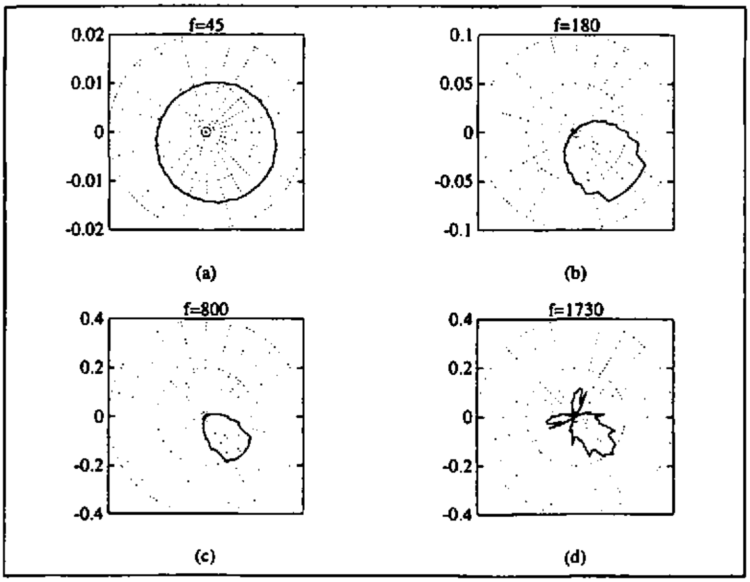

Figure 3. The power spectral density of $v_1(n)$, the sequence input to source number one of Figure 2 when plane waves producing a white noise sequence is recorded by the four sensors shown in Figure 2 and processed using the optimal filter matrix **H**. *The power spectral density is shown as a polar plot as a function of θ on a linear scale at 45Hz (a), 180Hz (b), 800Hz (c), and 1730Hz (d)*

4. EXPERIMENTAL RESULTS

Some experiments have been made using the geometrical arrangement shown in Figure 2. The 16 impulse responses that define the matrix **C** of acoustic transfer functions were measured in an anechoic chamber using a YDAP for the signal processing. The YDAP has generously been made available to us by Yamaha in Japan, it is a prototype, and it has a very high performance specification. The sample rate is 48kHz, and the measured impulse responses can contain as many as 16000 coefficients.

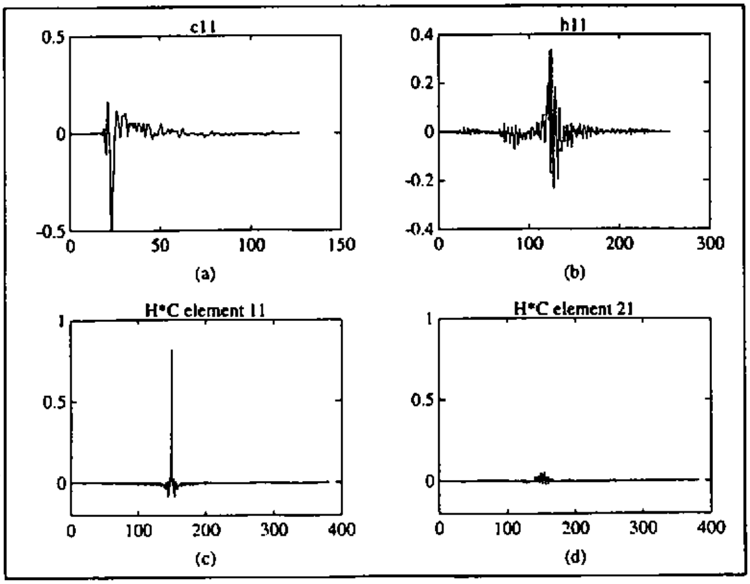

Figure 4. Some experimental results. (a) $c_{11}(n)$, (b) $h_{11}(n)$, (c) calculated response of element 1,1 of **H*C**, *and (d) calculated response of element 2,1 of* **H*C**

Figure 4a shows the measured impulse response from loudspeaker one to microphone one, $c_{11}(n)$. This impulse response was found by first windowing the original impulse response and then decimating it by a factor of four in order to make the adaptive deconvolution as quick possible. Thus, each element of **C** contains 128 coefficients. The 16 elements of the optimal filter matrix **H** were then found by running the adaptive LMS algorithm on a SUN workstation. Each element of **H** was chosen to contain 256 coefficients, and the modelling delay was 150 samples. Figure 4b shows the first element of the first row of **H**, $h_{11}(n)$. Note that most of the energy is concentrated near the centre of h_{11}, and that the response has almost decayed away at both the start and the end of the filter. This indicates a good choice of modelling delay and filter length. Two examples of the resulting output from the system with the optimal filters implemented are shown in Figures 4c and 4d. These responses are calculated by direct convolution of **C** with **H**, they are not measured. Figure 4c shows the impulse response from track one to microphone one, and Figure 4d shows the impulse response from track one, to microphone two. Ideally, Figure 4c should show a single digital impulse positioned at time sample number 150, which corresponds to the modelling delay, since we want to reproduce exactly the signal recorded on track one at the position of microphone one. Similarly, Figure 4d should ideally show a signal which is identically zero, because this element represents the "cross-talk" between track one and the position of microphone two. It is seen that the responses of the system with the optimal filters implemented are very close to ideal. However, more experiments are necessary in order to confirm the results of the computer simulations shown in Figure 3.

5. CONCLUSIONS

The work described in this paper is a preliminary attempt to evaluate the use of standard filter design techniques for the processing of recorded signals with the objective of reproducing the direction of arrival information in a restricted region of a listening space. The technique described shows some promise of success in preliminary computer simulations and in a simple experiment.

6. REFERENCES

[1] PA Nelson and SJ Elliott 1992 Active Control of Sound, London; Academic Press.
[2] PA Nelson 1993 Proc. Inst. Acoustics Spring Conference (Southampton).
[3] SJ Elliott and PA Nelson 1985 Electronics Letters 21, 979-981.
[4] SJ Elliott, IM Stothers and PA Nelson 1987 IEEE Trans. ASSP-35, 1423-1434