A FAST SUBBAND ROOM RESPONSE SIMULATOR

Firas Jabloun and Benoit Champagne

McGill University
Electrical and Computer Engineering Department
Montreal, Canada, H3A 2A7
firas@tsp.ece.mcgill.ca
champagne@ece.mcgill.ca

ABSTRACT

In this paper we present a subband scheme to decrease the computational complexity of simulating room responses to acoustic signals, important for instance in microphone array systems. Besides, the new method offers added flexibility to the well known image method by allowing to choose the reflection coefficients of every frequency subband independently of each others. The efficiency of this method is tested experimentally.

1. INTRODUCTION

Multi-microphone systems are increasingly gaining popularity in a variety of audio applications in adverse acoustic environments. Some of these applications are hearing aids, hands-free telephone communications and audio-conferencing. Multimicrophone systems need to reduce noise and suppress reverberation without introducing any noticeable distortion to the direct path signal.

To be able to design and evaluate such systems, it is helpful to have a fast way to digitally simulate the effect of a reverberant room on the acoustic signals at the input of the microphone array. In 1979, Allen and Berkley [1], introduced the image method to simulate the discrete-time room impulse response (IR). They represented the contribution of every image to the total effect, as an impulse shifted to the discrete time instant closest to the actual arrival time.

Unfortunately, the image method does not precisely estimate the echo arrival times because they are usually not multiples of the sampling period [6]. Hence it fails to give a good estimate to inter-microphone phase, which is essential for multi-microphone simulations. To solve this problem, Peterson suggests to distribute the arriving echo over several samples according to a low pass response function centered at the actual echo arrival time [6].

Usually the computation cost of the convolution operation needed to calculate the IR between two points in a reverberant room, is very large. Precisely, L multiplications per input sample, where L is the length of the impulse response, are required. Unfortunately, because of the room acoustic properties, L is generally a large number. This computational burden increases in the multi-microphone case because the costly convolution is repeated with all the microphones.

In this paper we present a fast subband room simulator (SRS) which significantly reduces the number of required multiplications. This is achieved using a subband scheme where the input signal is divided into K channels and every subsignal is convolved with a subband impulse response (SIR) at a sampling rate reduced by a factor $M \leq K$.

In addition to the complexity savings, our method provides much more simulation flexibility by allowing to adjust the acoustic parameters of every subband independently of the others. This is interesting because the absorption properties of rooms are known to vary in different frequency bands [4].

This paper is organized as follows, in Section 2 the overall subband system is described and computational savings are quantified. In Section 3, we describe two ways to calculate the SIR's and finally, in section 4, experimental results that show the efficiency of the SRS are reported.

2. ROOM SIMULATION USING SUBBAND FILTERING

Subband filtering has been used efficiently in various applications in the field of signal processing especially in speech coding and adaptive filtering. It owes its popularity to the frequency flexibility and the computational savings that it offers. In this paper, we exploit these two properties to solve the problem of computational complexity and to provide more dimensions of

0-7803-6293-4/00/$10.00

freedom to the currently used image methods.

2.1. Uniform DFT filter banks

In the proposed SRS, a uniform DFT filter bank is used to realize the subband analysis and synthesis. In this approach, the input signal is divided into K adjacent subbands by a bank of complex demodulators whose outputs are lowpass filtered by the antialiasing filter $f(n)$ and then downsampled by M to a lower sampling rate. After performing any desired modification on the subband signals, they are upsampled to the initial rate, lowpass filtered by the anti-imaging filter $g(n)$ and modulated back to their original spectral position. The output is finally obtained by summing the sub-signals. A critical design issue here is the choice of the downsampling factor M. In many practical implementations oversampling, that is $M < K$, is used. It is indeed one of the simplest and most efficient ways to reduce the subband aliasing when modifications are made to the subband signals [3]. While other methods could be used for the design and realization of oversampled DFT filter bank, we use the approach of [5], which was proved useful in the context of acoustic echo cancellation. A detailed analysis of this filter bank design is found in [5]. However for convenience of our upcoming derivations, we briefly summarize it here.

The z-transform $X_k(z)$ of the subband signals $x_k(m)$ (where $k = 0 \ldots K-1$ and m is the discrete time index at low sampling rate) is written as

$$X_k(z) = \frac{1}{M}\sum_{m=0}^{M-1} F(z^{1/M}W_M^{-m})X(z^{1/M}W_M^{-m}W_K^{k}) \tag{1}$$

where $W_K = \exp(j2\pi/K)$ and $W_M = \exp(j2\pi/M)$. $F(z)$ and $G(z)$ are the z-transforms of $f(n)$ and $g(n)$ respectively. We have the synthesizer output $y(n)$ in the z-domain

$$Y(z) = \frac{1}{M}\sum_{m=0}^{M-1} T_m(z)X(zW_M^{-m}) \tag{2}$$

The approach of [5] uses a special set of modulators (W_K^{-kn} for analysis and $W_K^{k(n+1)}$ for synthesis) which is slightly different from the conventional DFT filter bank. So assuming the sub-signals are not modified,

$$T_m(z) = \sum_{k=0}^{K-1} W_K^{k}F(zW_K^{-k}W_M^{-m})G(zW_K^{-k}) \tag{3}$$

Because $F(z)$ and $G(z)$ have the same cutoff frequency ($\omega_c = \pi/K$), and since $M < K$, $T_m(z) \approx 0$ for $m \neq 0$ [5]. So the aliasing components in (2) can be neglected and (2) reduces to

$$Y(z) = \frac{1}{M}T_0(z)X(z) \tag{4}$$

In order to have a constant group delay and hence reduce the phase distortion, $f(n)$ and $g(n)$ are chosen to be FIR filters of length N such that $g(n) = f(N-n-1)$. Now if $N = n_f K$, where n_f is an integer, then the frequency response in (4) becomes

$$T_0(e^{j\omega}) = e^{-j(N-1)\omega}\sum_{k=0}^{K-1} \mid F(e^{j\omega}W_K^{-k}) \mid^2 \tag{5}$$

Obviously the response of the filter bank has a linear phase. So by designing a filter bank such that the magnitude of $T_0(e^{j\omega})$ is equal to one, the total distortion will be a pure delay. In [5], the design of $f(n)$ to approximate these requirements is achieved by interpolation of 2-channel QMF filters (found in lookup tables [2]) by a factor of $K/2$.

For the implementation of the oversampled DFT filter bank, we use the *weighted-overlap-add* (WOA) structure because it allows flexibility in choosing M as any arbitrary integer. This approach needs in the order of $2(\log_2 K + n_f)K/M$ multiplications per input sample (MPIS), which is generally a very low cost.

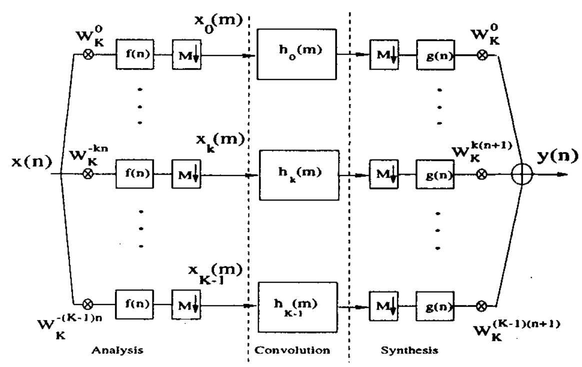

Figure 1: Block diagram of the SRS algorithm.

2.2. The Subband Room Simulator (SRS)

The subband analysis/synthesis system so designed can be used to perform the convolution with the room SIR at the low sampling rate. Our goal is to have an SRS with IR $\hat{h}(n)$ as close as possible to a reference full band IR $h(n)$ representing the desired transmission characteristics of the room. Figure 1 shows a block diagram of the SRS.

At the reduced sampling rate, the SIR's $h_k(m)$ have a length L/M where L is the length of $h(n)$. Consequently, L/M multiplications are needed every M samples for all the K channels. So, KL/M^2 MPIS are required to accomplish the convolution operations. With $h(n)$, on the other hand, the convolution needs L MPIS. Hence the total computational gain is

$$\frac{M^2}{K}\left[\frac{1}{1+\alpha(L)}\right] \tag{6}$$

where $\alpha(L) = 2M(\log_2 K + n_f)/L$. For most practical cases, L is very large so $\alpha(L)$ is usually a fraction between 0 and 1. Asymptotically, $\alpha(L) \to 0$ so the savings are in the order of M^2/K.

Actually since the subband signals are complex, the convolution requires twice as much multiplications as stated above, however this can be overcome using the symmetry properties of subband signals and avoid unnecessary repeated calculations. Namely, $X_k(m) = X^*_{K-k}(m)$ for $k = 0 \ldots K/2-1$. Moreover, according to the room properties, the lengths of the SIR's can be shortened , if the Energy Decay Curve (EDC) [4, 1] at every subband, is judged to have a steep decay.

3. CALCULATING THE SUBBAND IMPULSE RESPONSES

Calculating the SIR's is a critical design problem in our method. In this section we describe two ways to calculate $h_k(m)$. The first, which we will call the direct method, calculates the SIR's from a desired full band reference IR $h(n)$. The latter can be obtained from real measurements or can be a synthetic one calculated using the lowpass impulse method of [6], for instance. The second one, called the subband image method, uses the image method to calculate $h_k(m)$ in a subband scheme.

3.1. The Direct Method

In this section, we show how to calculate the SIR $h_k(m)$, so that the system in Figure 1, has an IR equal to a desired room response $h(n)$. So $T_m(z)$ in (3) becomes

$$T_m(z) = \sum_{k=0}^{K-1} W_K^k F(zW_K^{-k}W_M^{-m})G(zW_K^{-k})H_k(z^M W_K^{-kM}) \tag{7}$$

Our aim is to find an $H_k(z)$ which if upsampled and modulated (due to applying the synthesis bank to it) makes $H(z)$ appear in (7) as the transfer function of the whole system. One way to do that, is to feed $h(n)$ to an analysis bank so that

$$H_k(z) = \frac{1}{M}\sum_{m=0}^{M-1} \tilde{F}(z^{1/M}W_M^{-m})H(z^{1/M}W_M^{-m}W_K^k) \tag{8}$$

where $\tilde{F}(z)$ is the z-transform of $\tilde{f}(n)$ which is an FIR lowpass filter with cutoff frequency $\gamma\pi/K$. If $\gamma < K/M$ then we can substitute (8) in (7) and again argue that $T_m(z) \approx 0$ for $m \neq 0$, to have

$$Y(z) = \frac{1}{M^2}A(z)H(z)X(z) \tag{9}$$

where in the frequency domain $A(e^{j\omega})$ is

$$A(e^{j\omega}) = e^{-j(N-1)\omega}\sum_{k=0}^{K-1} |F(e^{j\omega}W_K^{-k})|^2 \tilde{F}(e^{j\omega}W_K^{-k}) \tag{10}$$

Our experiments show that if $\tilde{f}(n) = f(n)$, then the magnitude of $A(e^{j\omega})$ will have peaks at the boundaries of the frequency subbands which significantly distorts the final result. Therefore, $\tilde{f}(n)$ is chosen to have a slightly bigger bandwidth than $f(n)$ (that is $\gamma > 1$).

3.2. The Subband image method

Next, we derive the second method to calculate the SIR's. In the lowpass impulse method [1, 6], the response $h(n)$ of a microphone at location x to an impulse excitation at location x_0 can be expressed as follows

$$h(n) = \frac{c}{4\pi}\sum_r \frac{\beta_r}{\tau_r}\psi_r(n) \tag{11}$$

where c is the speed of sound, r is the index of the image at position x_r, $\tau_r = \| x_r - x \| /c$ is the echo arrival time of the image r, β_r is the corresponding composite reflection coefficient and $\psi_r(n)$ is a sampled version of a continuous lowpass filter centered at the echo arrival time τ_r. Now (8) written in the time domain becomes

$$h_k(m) = \sum_n \tilde{f}(mM-n)h(n)W_K^{-kn} \tag{12}$$

Then, substituting (11) in (12) and after changing the summation order we get

$$h_k(m) = \frac{c}{4\pi}\sum_r \frac{\beta_{r,k}}{\tau_r}\,\xi_{r,k}(m) \quad k = 0\ldots K-1 \tag{13}$$

where

$$\xi_{r,k}(m) = \sum_n \tilde{f}(mM-n)\psi_r(n)W_K^{-kn} \tag{14}$$

Note that a subscript k is added to the composite reflection coefficient β_r to indicate that with this formulation it is possible to assign different reflection properties to each subband. The way to implement (13) and (14) is to perform subband analysis to the lowpass function corresponding to every new image and update the subband responses $h_k(m)$ accordingly. Once the SIR's $h_k(m)$ are computed they are used as in Figure 1 to simulate a synthetic room IR;

4. EXPERIMENTAL RESULTS

In this section we describe the experiments made to test the performance of the new method and to compare its simulated room impulse response with that of the original image method. The room to be simulated has dimensions (15,10,4) with a microphone at positions (9,3.75,0.7) and a loudspeaker at (2,1.5,1.5). The dimensions are in meters and the origin of the coordinate system is at one of the lower corners of the room. The reflection coefficients are set to 0.9 for the walls and 0.7 for the floor and ceiling. The length of the impulse response is 128 msec that is $L = 2048$ samples at 16 KHz sampling rate. For the filter bank we use the

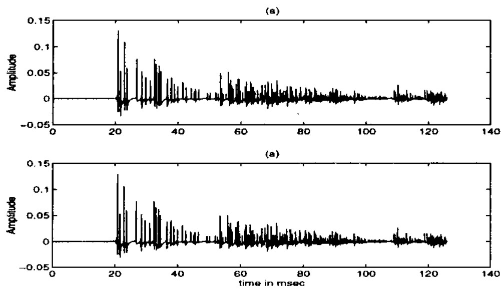

Figure 2: Image method impulse response $h(n)$ (up) SRS impulse response $\hat{h}(n+d)$ (down).

following parameters, $K = 32$, $M = 24$, $N = 256$.The prototype filter $f(n)$ is obtained via interpolation (see [5]) from the 2-channel QMF filter 16A in [2] ($n_f = 8$). The second filter $\tilde{f}(n)$ is designed by windowing (using a Hamming window) with $\gamma = 1.2$. With these parameters $\alpha(L)$ in (6) is 0.3 leading to a gain of about 13 times.

The IR $h(n)$ computed with the image method [1, 6], and the SRS IR $\hat{h}(n+d)$ (where d is the group delay of $A(e^{j\omega})$) computed with the direct method of Section 3.1, are shown in Figure 2. Figure 3 illustrates the magnitude squared of the Fourier transform $E(e^{j\omega})$ of the error function $e(n) = h(n) - \hat{h}(n+d)$. The arithmetic mean of the error squared is as low as -40 dB.

Additional experiments were made with audio data of speech, music and percussion recordings . A few persons were asked to listen to the recordings after adding reverberation to them using full band convolution with $h(n)$ and using the SRS (the reverberation time was 256 msec). To have some psychological effect, these people were indirectly given the impression that their ability to detect some kind of difference between two audio recordings is being tested. However, they disoppointedly reported their inability to perceive any difference.

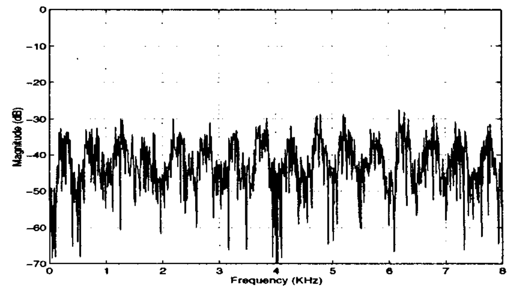

Figure 3: Magnitude Squared of the error $|E(e^{j\omega})|^2$

5. CONCLUSION

In this paper, a fast Subband Room Simulator (SRS) is presented. Basically the method uses a subband filter bank to perform convolution operations at a reduced sampling rate, hence reducing the computational complexity. This reduction facilitates the evaluation of algorithms designed for the increasingly growing research area of multi-microphone systems. Furthermore the SRS can be implemented in a way that offers more flexibility in choosing the room acoustic parameters. Our experiments show that no difference can be perceived between reverberation added using the traditional image method or using the SRS.

6. REFERENCES

[1] J. B. Allen and D. A. Berkley, *Image method for efficiently simulating small-room acoustics*, J. Acoust. Soc. Amer. **65** (1979), no. 4, 943–950.

[2] R. E. Crochiere and L. R. Rabiner, *Multirate digital signal processing*, Prentice-Hall, Englewood Cliffs, NJ, 1983.

[3] A. Gilloire and M. Vetterli, *Adaptive filtering in subbands with critical sampling: analysis, experiments and application to acoustic echo cancellation*, IEEE Trans. Signal Processing **40**, 1862–1875, Aug. 1992.

[4] H. Kuttruff, *Room acoustics*, 3rd Edn. Elsevier, London, 1991.

[5] Q. Liu, B. Champagne, and D. K. C. Ho, *Simple design of oversampled uniform DFT filter banks with applications to subband acoustic echo cancellation*, Proc. IWAENC-97 (1997), p. 132.

[6] P. M. Peterson, *Simulating the response of multiple microphones to a single acoustic source in a reverberant room*, J. Acoust. Soc. Amer. **80**, no. 5, 1527–1529, Nov. 1986.