FAST PERCEPTUAL CONVOLUTION FOR ROOM REVERBERATION

Wen-Chieh Lee and Chi-Min Liu

Dept. of Computer Science and Information Engineering
National Chiao Tung University, Taiwan
wjlee@csie.nctu.edu.tw

Chung-Han Yang and Jiun-In Guo

Dept. of Computer Science and Information Engineering
National Chung Cheng University, Taiwan
chyang@csie.nctu.edu.tw

ABSTRACT

The FIR-based reverberators, which convolve the input sequence with an impulse response modelling the concert hall, have better quality compared to the IIR-based approach. However, the high computational complexity of the FIR-based reverberators limits the applicability to most cost-oriented system. This paper introduces a method that uses perceptual criterion to reduce the complexity of convolution methods for reverberation. Also, an objective measurement criterion is introduced to check the perceptual difference from the reduction. The result has shown that the length of impulse response can be cut off by 60% without affecting the perceptual reverberation quality. The method is well integrated into the existing FFT-based approach is have around 30% speed-up.

1. INTRODUCTION

Artificial reverberators have been used to add reverberation to studio recording in the music and film industry, or to modify the acoustic of a listening room. There have been basically two approaches to design reverberators. The first approach is based on the IIR (Infinite Impulse Response)-recursive networks such as comb filters, all-pass filters. A variety of algorithms [8][9][10] have been proposed since the work of Schroeder [5][6]. The IIR-based network has the merit in low complexity, but is often difficult to eliminate unnatural resonances. On the other hand, the FIR (Finite Impulse Response)-based reverberators, which convolve the input sequence with an impulse response modeling the concert hall, will be free from the unnatural resonances. However, the high computational complexity due to the long FIR length leads to another concern in real-time applications. For the two seconds of impulse response, the length will be 88,200 samples in terms of 44,100Hz sampling rate. Using direct convolution to implement the reverberation indicates the 88,200 multiplications for each sample, or 7.8G multiplications per second for stereo audio.

A lots of researches [1][3][15] have been developed to reduce the complexity of FIR-based reverberators. Among them, the FFT-based methods can significantly reduce the complexity. This paper proposes a new idea in reducing complexity by combining the perceptual phenomenon with the FFT based method called fast perceptual convolution. Besides, for having an effective quality measurement on the fast perceptual convolution, we examine the quality through an objective criterion which compares the perceptual difference between the tracks processed through the non-reduced FIR and the perceptually-reduced FIR. The result has shown a 30% improvement without affecting the perceptual reverberation quality.

2. BLOCK CONVOLUTION

FIR-based reverberators are implemented by convolution methods. This section will introduce the operation of convolution and the block convolution method for the reverberators.

Because we need to process segmented input signal, methods to recombine the processed segments into final signal are needed. There have been two approaches: overlap-and-save [13] method and overlap-and-add [14] method. The convolution between input signal $x[n]$ and impulse response $h[n]$ of length L is expressed as

$$y[n] = x[n] * h[n] = \sum_{k=0}^{L-1} x[n-k]h[k] \qquad (1)$$

The overlap-and-add method adopts non-overlapped input segments to calculate overlapped output segments. To extend the overlap-and-add approach to segmented impulse response, let the input signals $x[n]$ and impulse response $h[n]$ are segmented as a sum of shifted finite-length segments of length N; i.e.,

$$x[n] = \sum_{r=0}^{\infty} x_r[n-rN] \qquad (2)$$

and

$$h[n] = \sum_{s=0}^{M-1} h_s[n-sN] \qquad (3)$$

where M is the smallest integer larger than L divided by N, i.e. $M = \left\lceil \frac{L}{N} \right\rceil$

$$x_r[n] = \begin{cases} x[n+rN], & 0 \le n \le N-1 \\ 0, & \text{otherwise} \end{cases} \qquad (4)$$

and

$$h_s[n] = \begin{cases} h[n+sN], & 0 \le n \le N-1 \\ 0, & \text{otherwise} \end{cases} \qquad (5)$$

Substituting (2) and (3) into (1) yields

$$y[n] = \left\{ \sum_{r=0}^{\infty} x_r[n-rN] \right\} * \left\{ \sum_{s=0}^{M-1} h_s[n-sN] \right\} \qquad (6)$$

Because convolution is linear time-invariant, it follows that

$$\begin{aligned} y[n] &= \sum_{r=0}^{\infty}\sum_{s=0}^{M-1} x_r[n-rN] * h_s[n-sN] \\ &= \sum_{r=0}^{\infty}\sum_{s=0}^{M-1} y_{r,s}[n-rN-sN] \end{aligned} \tag{7}$$

where

$$y_{r,s}[n] = x_r[n] * h_s[n] \quad \text{for } 0 \le n < 2N-1 \tag{8}$$

In overlap-and-add method, the convolution of each pair of small blocks can be transformed to DFT domain and perform multiplications on DFT domain. Since the complexity of the specific sizes of DFT can be reduced from O(N^2) to O($N\log N$) by FFT algorithms, these algorithms can perform the convolution with significant speed improvement.

3. FAST PERCEPTUAL CONVOLUTION

The fast perceptual convolution we proposed is to reduce the computational complexity required by FIR-based reverberators. There have been research [15] trying to reduce the complexity by modifying the impulse response through perceptual criterion. However, that approach adopts the time-domain method that can not be jointly combined with the FFT-based approach to have complexity merit. The fast perceptual convolution proposed in this paper reduces the multiplications needed in frequency-domain and can be integrated well with the FFT-based convolution methods to have lower complexity.

3.1. Block Convolution Performed through FFT

We discussed that the linear convolution of a long impulse response, we can separate both input signal $x[n]$ and impulse response $h[n]$ into blocks. The convolution for each pair of input signal block $x_r[n]$ and impulse response block $h_s[n]$ can be implemented with the FFT with $2N-1$ points. We adopt for complexity evaluation based on radix-2 FFT and $2N$-point FFT instead of ($2N-1$)-point FFT.

Since the convolution in time domain is equivalent to the multiplication in frequency domain, (8) can be written as

$$Y_{r,s}[k] = X_r[k]H_s[k], \quad \text{for } 0 \le k < 2N \tag{9}$$

where $Y_{r,s}[k]$, $X_r[k]$ and $H_s[k]$ are the $2N$-point FFT of $y_{r,s}[n]$, $x_r[n]$ and $h_s[n]$, respectively. Let $p=r+s$. (7) is rewritten as

$$\begin{aligned} y[n] &= \sum_{p=s}^{\infty}\sum_{s=0}^{M-1} y_{p-s,s}[n-pN] \\ &= \sum_{p=s}^{\infty}\sum_{s=0}^{M-1} x_{p-s}[n-(p-s)N] * h_s[n-sN] \end{aligned} \tag{10}$$

Define

$$\begin{aligned} y_p[n] &= \sum_{s=0}^{M-1} y_{p-s,s}[n-pN] \\ &= \sum_{s=0}^{M-1} x_{p-s}[n-(p-s)N] * h_s[n-sN] \end{aligned} \tag{11}$$

Hence,

$$y[n] = \sum_{p=s}^{\infty} y_p[n] \tag{12}$$

The nonzero values of $y_p[n]$ are only in the time interval [pN, $pN+2N-2$]. Let $n' = n - pN$, we have

$$y_p[n' + pN] = \sum_{s=0}^{M-1} y_{p-s,s}[n'] \tag{13}$$

Performing $2N$-point FFT on (11) within the nonzero interval [0, $2N-1$] leads to

$$\begin{aligned} Y_p[k] &= \sum_{s=0}^{M-1} Y_{p-s,s}[k] \\ &= \sum_{s=0}^{M-1} X_{p-s}[k]H_s[k], \quad \text{for } 0 \le k < 2N-1 \end{aligned} \tag{14}$$

The fast convolution is summarized as follows:

Step 1: Store the FFT data of the segmented impulse response, $H_s[k]$.

Step 2: Execute $2N$-FFT on the segmented input signals to obtain $X_r[k]$.

Step 3: Multiply and add the two FFT data according to (14). The number of multiplications and additions is both $M+M/N$ for each input sample.

Step 4: Perform inverse FFT to have the segmented data $y_p[n]$.

Step 5: Overlap and add all the segmented $y_p[n]$ to have the final $y[n]$ according to (14). The overlapping factor is 1 and hence has the complexity one.

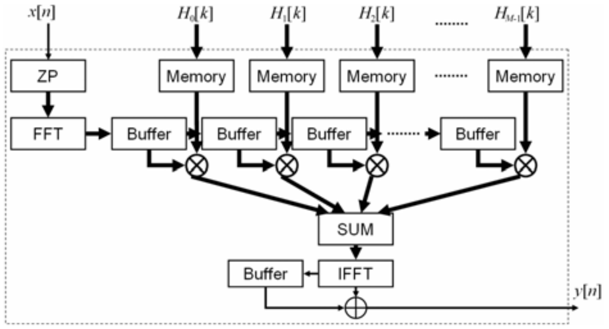

Figure 1: *Block diagram of fast convolution*

The block diagram of the fast convolution is illustrated in *Figure 1*. The complexity of multiplications in fast convolution is $2\mathrm{FFT}(2N)/N+M+M/N$.

3.2. Perceptual Convolution

The threshold in quiet is the threshold to characterize the minimum amount of energy needed in human hearing system in a noiseless environment. One well known threshold is the one made by Painter and Spanias [2]. The threshold provides the potential to sparse the $H_s[k]$ and hence reduces the complexity.

Consider (14), the output signal $Y_p[k]$ will not be perceptible if the energy is lower than the threshold in quiet. That is

$$\left|Y_p[k]\right| \le Th[k]. \tag{15}$$

where $Th[k]$ is the threshold in quiet for a frequency k. Substituting (14) to (15) leads to

$$\left|\sum_{s=0}^{M-1} X_{p-s}[k]H_s[k]\right| \le Th[k], \text{ for } 0 \le k < 2N-1 \tag{16}$$

Assume the signal magnitude is lower than ρ, (16) is reduced to

$$\left|\sum_{s=0}^{M-1} X_{p-s}[k]H_s[k]\right| \le \rho\left|\sum_{s=0}^{M-1} H_s[k]\right| \le Th[k], \quad \text{for } 0\le k<2N-1 \tag{17}$$

The sufficient condition for the above inequality on $|H_s[k]|$ is

$$\left|H_s[k]\right| \le \frac{Th[k]}{M\rho}, \quad \text{for } 0\le k<2N-1 \tag{18}$$

In other words, we can directly truncate the small values of $|H_s[k]|$ into zeros to reduce the complexity according to (18). *Figure 2* illustrates the spectrum for all the segmented impulse responses. The higher frequency part will decay faster than lower frequency part. *Figure 4* is the resulted spectrum after the sparse process according to (18). The block diagram of fast perceptual convolution is shown in *Figure 3*.

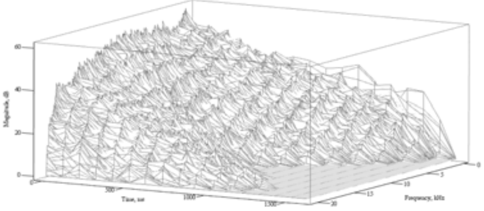

Figure 2: *Spectrum of the impulse response recorded from St. John Lutheran Church*

Table 1: *Percentage of eliminated frequency domain multiplications of each impulse response when the block size is set to 4,096*

Impulse Response	St John Lutheran	Foellinger Great Hall	Bethel Church	Meyerson Concert Hall
Eliminated Percentage	61.47%	59.33%	60.99%	62.79%

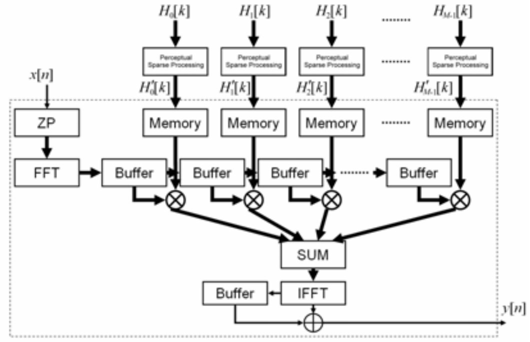

Figure 3: *Block diagram of fast perceptual convolution*

Table 1 shows the percentage of eliminated frequency domain multiplications of 4 different impulse responses. For those impulse responses, we can eliminate more than 50% of multiplications in frequency domain. For some blocks, we can remove the multiplications for the whole block. *Figure 4* shows the same impulse response as that in *Figure 2* after removing ignored frequencies.

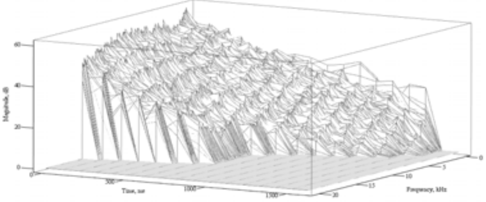

Figure 4: *Spectrum of the impulse response of St. John Lutheran Church after applying the perceptual threshold*

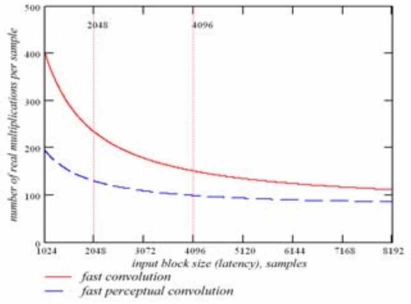

Figure 5: *Comparison of fast perceptual convolution and fast convolution when the length impulse response is 2 seconds*

3.3. Results

Assuming that we can remove 60% of multiplications in frequency domain, we can calculate the number of multiplications needed for fast perceptual convolution by modifying the complexity from fast convolution as illustrated in *Figure 5*. From the result, the fast perceptual convolution requires about 98 real multiplications per sample to convolve with 88,200 samples of impulse response.

To evaluate the improvement in real-time systems, we built an experiment application to help evaluate the test. The application uses two methods, the fast perceptual convolution and fast convolution, to process some samples for comparison. The input block size is set to 4,096. And the test is to process single channel, 4,096×20,000 = 81,920,000 samples of input, which is about 30 minutes of samples with 44,100Hz sampling rate. The test is run on a PC with 1GHz Pentium *!!!*. The result is listed in *Table 2*.

Table 2 shows that the fast perceptual convolution can reduce about 30% complexity as compared with the fast convolution in real applications.

Table 2: *Comparison of fast perceptual convolution and fast convolution*

Time in ms	St John Lutheran	Foellinger Great Hall	Bethel Church	Meyerson Concert Hall
Fast convolution	89469	88027	84692	82549
Fast perceptual convolution	59566	61057	58694	57032
Improved Ratio	33.42%	30.64%	30.70%	30.91%

4. OBJECTIVE MEASUREMENT

The fast perceptual convolution exploits the perceptual irrelevancy to develop fast convolution. This section considers the objective measure to check the irrelevancy. The Objective Difference Grade (ODG) which is suggested by Recommendation ITU-R BS.1387 [4] is introduced to the measurement. The ODG is the output variable from the objective measurement method. The values of ODG range from 0 to -4, where value 0 corresponds to an imperceptible impairment and value -4 to the impairment judged as very annoying. The system has included a subtle perceptual model and has been widely used in audio compression community to detect the perceptual difference between two tracks.

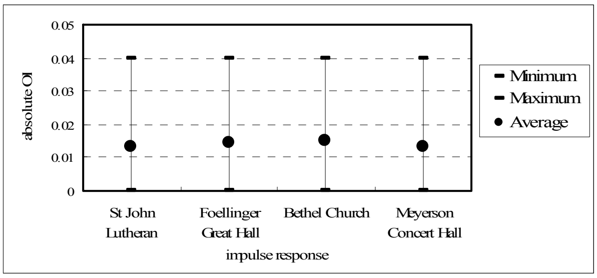

Figure 6: *Distribution of ODGs for fast perceptual convolution with different impulse responses*

To measure the reverberation quality of fast perceptual convolution, we use ODG to compare the reverberation generated by generic convolution methods and fast perceptual convolution. We did the test on 60 different samples for 4 different Hall reverberations. The result is summarized in *Figure 6*.

As shown in *Figure 6*, all the mean ODG values for four impulse responses are about 0.015 and the maximum ODG are all 0.04. These results indicate that the differences between the outputs of the proposed method and the original method are not perceptually noticeable. Hence, the fast perceptual convolution can reduce the length of the impulse response by 60% and has a speedup 30% on the reference design system over the FFT-based convolution without scarifying the reverberation quality.

5. CONCLUSIONS

This paper has proposed a fast perceptual convolution method which can cut off 60% impulse response without scarifying the reverberation quality. Since the fast perceptual convolution needs less than 100 real multiplications per input sample to convolve with the reverberation length as high as two seconds. This complexity is very close to that of the methods with IIR-based approach. In other words, the fast convolution seems to be a good tradeoff between the IIR-based and FIR-based approaches from the aspect of the complexity and the reverberation quality.

In addition to the merits in computational complexity, fast perceptual convolution reduces the memory requirement. Using proposed method can save about 60% of memory storage. This feature also benefits the realization of the proposed perceptual approach through DSP processors.

6. REFERENCES

[1] D. S. McGrath, "Method and Apparatus for Filtering an Electronic Environment with Improved Accuracy and Efficiency and Short Flow-Through Delay", US Patent 5,502,747.

[2] T. Painter and A. Spanias, "Perceptual Coding of Digital Audio", *Proceeding of The IEEE*, Vol. 88. No. 4, April 2000.

[3] W. G. Gardner, "Efficient Convolution without Input-Output Delay", *J. Audio Eng. Soc.*, vol. 43, no. 3, pp. 127-136, March 1995.

[4] ITU Radiocommunication Study Group 6, "DRAFT REVISION TO RECOMMENDDATION ITU-R BS.1387 - Method for objective measurements of perceived audio quality".

[5] M. R. Schroeder and B. F. Logan, "Colorless Artificial Reverberation", *J. Audio Eng. Soc.*, vol. 9, no. 3, pp. 192-197, July 1961.

[6] M. R. Schroeder, "Natural Sounding Artificial Reverberation", *J. Audio Eng. Soc.*, vol. 10, no. 3, pp. 219-223, July 1962.

[7] J. A. Moorer, "About This Reverberation Business", *Computer Music Journal*, vol. 3, no. 2, pp. 13-28, June 1979.

[8] J. M. Jot and A. Chaigne, "Digital delay networks for designing artificial reverberators", in *Proc. 90th Conv. Audio Eng. Soc.*, February, 1991, preprint 3030.

[9] W. G. Gardner, "The virtual acoustic room", MS paper, MIT Media Lab, 1992. http://alindsay.www.media.mit.edu/papers.html.

[10] J. Dattorro, "Effect Design Part 1: Reverberator and Other Filters", *J. Audio Eng. Soc.*, vol. 45, pp. 660-684, September 1997.

[11] H. V. Sorensen, D. L. Jones, M. T. Heideman, and C. S. Burrus, "Real-Valued Fast Fourier Transform Algorithms", *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-35, June 1987.

[12] W. C. Sabine, "Reverberation", in Lindsay, R.B., editor, Acoustic: Historical and Philosophical Development, Stroudsburg, PA: Dowden Hutchinson, and Ross, 1972. Originally published in 1990.

[13] T. G. Stockham, Jr., "High-Speed Convolution and Correlation", in *Spring Joint Computer Conf.*, *AFIPS Conf. Proc.*, vol. 28, pp. 229-233, 1966; reprinted in Digital Signal Processing, Selected Reprints, L. R. Rabiner and C. M. Rader, Eds. (IEEE Press, New York, 1972).

[14] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1975

[15] K. Iida, K. Mizushima, Y. Takagi, and T. Suguta, "A New Method of Generating Artificial Reverberant Sound", *99th AES Convention 1995*, October 6-9, 1995.

[16] A. Torger and A. Farina, "Real-Time Partitioned Convolution for Ambiophonics Surround Sound", *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics 2001*, October 2001.

[17] J. S. Soo, K. K. Pang, "Multidelay block frequency adaptive filter", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. ASSP-38, No. 2, February, 1990.