Effects Processing on Audio Subband Data

Scott N. Levine
Stanford University
Center for Computer Research in Music and Acoustics (CCRMA)
scottl@ccrma.stanford.edu
http://www-ccrma.stanford.edu

Abstract

This paper will show that computing standard audio effects algorithms such as reverberation, echo, chorus and flange will use less memory and computations when performing the operations on the critically sampled subband data than on the fullrate time domain signal. Not only can effects in the subband domain be obtained to sound close to the effects in the time domain, but new types of effects are now possible because different effects can be placed in separate subbands. The MPEG Audio filter bank, which also splits the audio into subbands, is used in this discussion to show that standard MPEG audio decoders could easily be augmented to include effects processors as well.

1 Introduction

Since the MPEG Audio (layers I,II) standard has become a prominent data compression algorithm in the consumer multimedia market, it exists on many platforms ranging from personal computers to workstations to video games. For most of these applications, some sort of audio post-processing is desired by the end-user. A simple example could be the addition of artificial reverberation for watching a movie in a home theater.

These post-processing algorithms are widely available today, but usually require special hardware for its computation and large amounts of memory. It will be shown that these post-processing algorithms can be performed on the audio subband data itself present inside the MPEG decoding standards [Rothweiler, 1983]. By computing the effects while performing the MPEG decoding, the need for external effects processors is eliminated. Since effects are computed separately for each subband, custom tailored effects can now be placed on different regions of frequencies.

2 Standard Effects Processing Algorithms

Before delving into the benefits of computing effects in a multirate domain, some of the simplified, standard effects algorithms will be quickly explained. For a thorough explanation on these effects, see [Orfanidis, 1996].

2.1 Echo

The simplest effects example would be an echo simulation. An echo algorithm models discrete acoustic reflections from surfaces far away from the listener. If the surfaces are too close in space and the reflections are too close together in time, the ear will not recognize them as discrete echoes, but rather as a reverberation (which will be discussed later). To model an echo, all that is needed is a feedback comb filter with D uniformly spaced poles, as seen in figure 1.

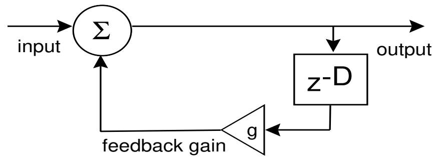

Figure 1: *Echo effect*

2.2 Flange

With digital technology, flanging translates to a delay line whose feedforward read pointer from the delay line to the output is sinusoidally modulated at around 0.2 Hz, as seen in figure 2. To make the structure general, the feedback read pointer from the delay line to the input summation is also shown as modulated. Traditional time domain flangers use no feedback modulation (i.e. a modulation of 0 Hz) due to its inherent pitch modulations. It will be shown later that this feedback modulation frequency is important when designing multiband flangers.

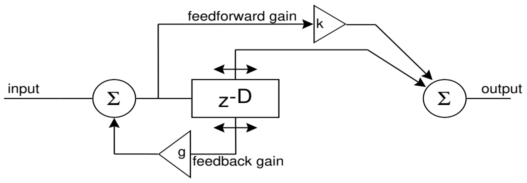

Figure 2: *Flange effect*

2.3 Chorus

The chorus effect is meant to make a single vocalist or instrumentalist sound like a whole group or section performing. When a group of people perform together, they are not all signing or playing at the exact same time, but are slightly nonsynchronous. The chorus effect tries to model the time variations between players with several modulated delay lines. Its topology is much like the flanger, except there is no feedback gain, and there are three parallel modulated delay lines instead of one, as shown in figure 3. The delay lines are longer than the flanger, around 25-30 milliseconds and the modulating frequency is still under 1 Hz.

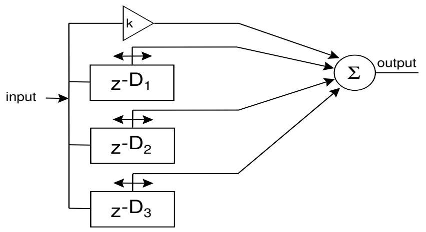

Figure 3: *Chorus effect*

2.4 Reverberation

A huge amount of papers have been published in the area of artificial reverberation, but many of them point towards [Schroeder, 1962] as the pioneer. The basic problem is how to make efficient models of diffuse acoustic reverberation without having to convolve a signal with the impulse response of the room. This response could last for seconds, and would require an FIR filter with tens of thousands of taps. The basic building blocks of these types of artificial reverberators consist of parallel banks of feedback comb filters (FBC), along with a series chain of allpass (AP) filters. The earliest such structure was shown by Schroeder, as seen in figure 4.

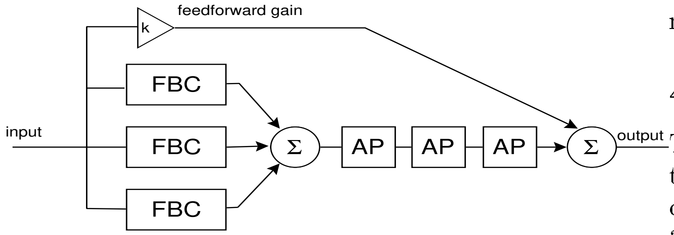

Figure 4: *Reverberation effect*

3 MPEG Filter Bank

When choosing a filter bank in which to implement frequency domain effects, the filter bank from the MPEG audio standards [Brandenburg and Stoll, 1994] proves to be a wise choice. First of all, the 32 channel, uniform cosine modulated pseudo quadrature mirror filter has high frequency selectivity. With a length of 512, the steep window response has their first sidelobes down over $100dB$. The magnitude response of this filter bank is shown in figure 5. Secondly, the computationally expensive matrix multiplication associated with the filter bank can be replaced with efficient, fast discrete cosine transform (DCT) algorithms.

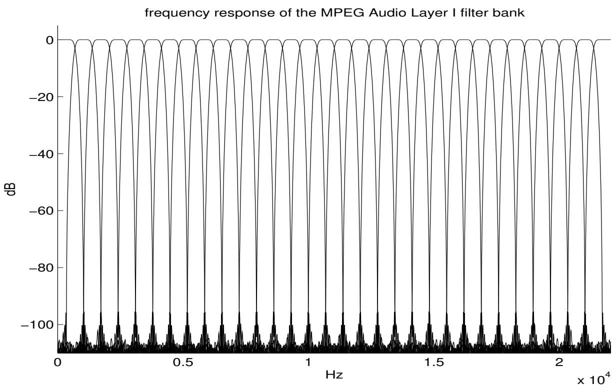

Figure 5: *Magnitude response*

Perhaps as important, the MPEG audio standard and its filter bank is already implemented on millions of platforms around the world. Therefore, the effects described in this paper could easily be inserted into the huge number of existing MPEG audio decoding systems. There would be no need to add other external hardware or software solutions to gain effects processing algorithms; the effects can be combined into the MPEG decoder itself.

4 Subband Effects Processing

This section will explain how effects are transformed into this MPEG filter bank scheme and remain sounding close to the effects computed on the time domain, fullrate signal. Then it will be shown that by altering these effects algorithms, one can gain sizable savings in computation and memory.

4.1 Simple Subband Effects

The first step to mapping an effects algorithm into the individual subbands is to reduce the length of any delay element in the original algorithm by 32, and multiply any modulation frequency by 32. These initial modifications are necessary because the subbands are critically downsampled by a factor of 32.

As a simple example of the effects transformation, a echo effect is desired with a delay line length of 300 msec. To replicate this effect exactly in the subband domain, first design another comb filter like the one shown in figure 1, with the same feedback scalars but with a $300msec./32 = 9.4msec$ delay line length. This altered comb filter is placed on every subband, as in figure 6. Then the processed subbands are run through the same MPEG synthesis filter bank, and the output will sound just like a standard, fullrate echo effect.

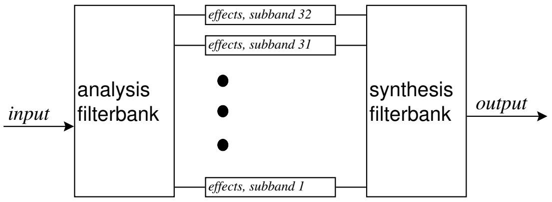

Figure 6: *Subband effects structure*

4.2 Customized Subband Effects

Because there are now 32 parallel, independent effects processors working over different bandwidths, we can now tailor the coefficients in each band separately. This gives the sound designer many more degrees of freedom than using a single fullrate effects algorithm.

4.2.1 The Multiband Flanger

With all these new degrees of freedom, we designed a frequency dependent flanger. Refer to figure 2 for the filter structure that will be placed in all subbands. First of all, we sinusoidally modulate both the feedforward and feedback pointers in all subbands. In a fullrate flanger, feedback modulation is usually avoided due to inherent pitch shifting qualities. But in this structure, any pitch shifting will be limited to a 687 Hz wide subband, so it is not as noticeable. Also, feedback modulation makes the multiband flanger sound closer to the fullrate version, for reasons we have yet to explain.

Secondly, we shorten the delay line lengths as the subband center frequency gets higher. In this system, we begin with 10 msec. delay lines at the lowest frequency subband, and linearly reduce their lengths to 3 msec at the 15th subband (at around 10 kHz). Above this frequency, the energy levels are so low, it is not worthwhile spending the memory and computation to process the signal. Thus, the subband data is passed through dry.

Also, more subtle changes can be heard if the gain magnitudes are decreased towards zero as the subbands get higher in frequency. In this way, the deepest harmonically spaced frequency notches are at lower frequencies only.

4.2.2 Multiband Reverberators

Most commercial reverberators are bandlimited to around 8kHz. This cutoff is due to the fact that the materials from most room interiors only reflect low frequencies. When looking at the magnitude responses of most rooms, the higher frequencies have quicker exponential decay rates than the lower frequencies. This translates to smaller delay line lengths for the higher frequencies. In [Schroeder, 1962], he defines the T_{60} time, which equals the time the reverberating sound level drops by 60 db. In a feedback comb filter, the delay line length D is related to the T_{60} and the feedback scalar $|g| < 1$ by the equation: $D = \frac{1}{3}T_{60} \cdot log(\frac{1}{|g|})$. For a fixed g, as the T_{60} decreases, so does D. With this subband structure in place, one can tailor these three preceding coefficients to resemble generic rooms.

For the included reverb sound example, only the first ten subbands were processed; the other 22 subbands were left dry and unprocessed. The first subband, which is lowest frequency band, had a $T_{60} = 1sec$. Subbands 2 through 10 had linearly decreasing T_{60} times (with correspondingly shorter delay lines), and subband 10 had a T_{60} time equal to one-third of the first subband. In this algorithm, the feedback factor g stays constant for all subbands. The allpass filters retained the same coefficients for all ten subbands, which used very little memory (under 5 msec.). In order to remove any periodicities in the frequency domain, some low level noise was added to these three parameters. A short-time Fourier transform of this reverberator's impulse response is shown in figure 7.

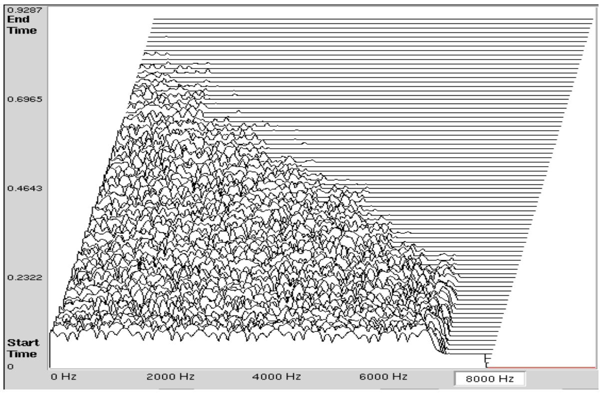

Figure 7: *Reverberator magnitude response*

4.3 Memory and Computational Savings

Assuming the cost of MPEG decoding is already provided for, the amount of calculations and words of memory required for effects processing of 32 critically sampled subbands are equal to that of a fullrate signal. This property is due to the fact that all the effects delay line lengths have been reduced by a factor of 32. The savings in memory come when the effects algorithms are individually altered for each subband. As shown in the previous two sections, only 10-15 of the total 32 subbands were processed. Additionally, the higher frequency subband effects algorithms used up to one-third the memory of the lower frequency ones. These multiband algorithms can shrink the amount of memory necessary by up to a factor of five, while still sounding like a convincing, fullrate commercial effect.

4.4 Using the Side Information

An optimization problem present is how to get an effects algorithm to sound the best with using a limited amount of memory and computations. Looking at the side information from the psychoacoustic masking function data provided by the MPEG encoder would be a good initial starting point for deciding which subbands are more perceptually relevant. If the MPEG encoder algorithm previously decided to allocate many bits to a certain subband, it would be worthwhile to allocate more memory and computation to the that particular subband's effect algorithm. This process is very efficient since it would be utilizing all the calculations previously performed in the MPEG encoder's complex psychoacoustic model [Brandenburg and Stoll, 1994]. Similarly, the encoder's transient detection information could help these algorithms adaptively change their coefficients in new and novel ways.

4.5 Layering Subband Effects

Instead of allocating all memory and computations for one effects algorithm, one could apply several effects over different (and possibly overlapping) frequency ranges. For example, the user could allocate all available memory for one good reverberator over all audible frequencies. Or, the user could split the memory between a slightly worse reverberator from $0 - 8kHz$, plus a flanger from $2 - 12kHz$ and a chorus from $4 - 10kHz$ by placing the effects in the corresponding subbands.

4.6 Alias Cancellation

If no processing were to take place between the analysis and synthesis filter banks in figure 6, then the any aliasing noise would remain below a $100db$ noise floor. The problem arises when the critically sampled subbands are processed. In [Schoenle, Fliege, and Zölzer, 1993], they avoided critically sampled systems and chose to implement artificial reverberation using an oversampled filter bank. By carrying around twice the amount of subband data, they avoid the problems of aliasing. To stay within the MPEG standards and reduce the memory requirements, their method is avoided.

In the experimental work with this MPEG multirate effects framework, no audible aliasing errors have been noticed. Even if there are aliasing problems, it could easily get perceptually buried under the alterations that the effect itself is creating. In addition, due to the structure of the MPEG filter bank, any aliasing by an effects algorithm in one band will be bandlimited by half a subband (343 Hz) on each side. Therefore, since the aliased regions are highly attenuated, it is most likely that their energy will be masked.

5 Conclusion

This paper has shown that memory and computation can be saved by moving post-processing audio effects such as reverb, chorus, flange and echo from the time domain to the subband domain. These subband domain audio signals are present in the current MPEG Audio compression standards. It was shown that effects processing can be calculated on the subbands while the audio is being MPEG decompressed. This combination of computations eliminates the need for specialized external effects processing hardware. Since the effects are calculated on the audio subbands separately, new and different effects can be placed on different regions of frequencies.

References

[Rothweiler, 1983] J. H. Rothweiler. *Polyphase Quadrature Filters - A New Subband coding Technique*, International Conference IEEE ASSP 1983, Boston, S.1280-1283.

[Orfanidis, 1996] S.J.Orfanidis. *Introduction to Signal Processing.* Prentice-Hall, 1996. pp. 355-383.

[Schroeder, 1962] M.R.Schroeder. *Natural Sounding Artificial Reverberation*, Journal of the AES, Vol. 10, No. 3, July 1962.

[Schoenle, Fliege, and Zölzer, 1993] M. Schoenle, N. Fliege, and U. Zölzer, 1993. *Parametric Approximation of Room Impulse Responses by Multirate Systems*, ICASSP 1993.

[Brandenburg and Stoll, 1994] K.Brandenburg and G. Stoll. *ISO-MPEG-1 Audio: A Generic Standard for Coding of High-Quality Digital Audio*, Journal of the Audio Engineering Society, Vol. 42, No. 10, October 1994.