

High-frequency compensation of low sample-rate audio files: A wavelet-based spectral excitation algorithm

Corey I. Cheng

coreyc@eecs.umich.edu
Department of Electrical Engineering, Systems Division
University of Michigan, Ann Arbor

Abstract

The present work documents the design and implementation of a wavelet-based excitation algorithm for low sample-rate audio files. This algorithm is an analysis/resynthesis method which converts a low sample-rate audio file to a higher sample-rate audio file, while inserting high frequency content in the wavelet domain. This paper presents relevant wavelet theory, along with the theory of the excitation algorithm and some sonic results. We find that the excitation algorithm works reasonably well for percussive, transient-oriented sounds, but fails for highly-resonant steady-state sounds, e.g., vocalic segments of speech.

1 Introduction and motivation

Although the high frequency range of a typical audio recording has little power compared with its lower frequencies, much information, such as localization and ambient cues, is contained in the higher frequencies. However, in any pragmatic physical system, high frequencies are usually quickly attenuated from a source due to dissipative forces such as friction and heat.

The many EQ-based mechanisms that have been introduced to accomplish high-frequency boosting and compensation are called "exciters" in the audio industry, and generally work by emphasizing a signal's existing but attenuated high-frequency content. However, if we wish to excite digital signals that exist at sampling rates lower than 44.1kHz, we face a far more acute problem, since only the frequencies from zero to the Nyquist frequency can be captured by the digitized audio file. Therefore, the high frequencies for which we might wish to compensate might not even be present in the original signal, in which case we must *guess* this high frequency content from the existing lower frequency content.

Wavelets are well suited to this predictive task. Since a dyadic wavelet analysis/resynthesis is an octave-band decomposition that can theoretically be done in linear time, an entire octave of high frequency information can be extrapolated from existing lower frequencies more efficiently than a corresponding Fourier technique. Furthermore, as is shown below, an exponential decay property of wavelet coefficients allows us to extrapolate high frequencies from a low sample-rate audio file in a straightforward manner [14].

2 Relevant wavelet theory

Wavelet analysis is one of many generalized time-frequency methods which serves to describe a signal's frequency content at a certain point in time. As opposed to Fourier analysis, however, wavelet analysis divides the time-frequency plane into non-uniform regions which are characteristic of an octave-band decomposition. In addition, each region of the time-frequency plane is associated with a time-domain signal called a wavelet. These wavelets form a basis for the analyzed signal, and, when added together in prescribed amounts, can perfectly reconstruct certain classes of signals. An interesting feature of wavelet analysis is that all of these time domain signals have the same general shape, and that in fact, they only differ by compressions and expansions by powers of two.

A level- n *wavelet analysis* produces n levels of *wavelet coefficients* and a *final average coefficient*, which serve to summarize in the wavelet domain the frequency content of the signal at a certain time. As shown in Figure 1, a higher-level wavelet coefficient corresponds to higher frequency content, wider frequency range, and a smaller time interval.

The forward and inverse dyadic wavelet transform of a window of samples can be implemented using a set of upsamplers, downsamplers, and recursive two-channel digital filter banks. The coefficients of each of the filters in the filterbank are directly related to both the time-domain shape of the wavelets, as well as the frequency response of the wavelet analysis. For the forward wavelet transform, the output of the filter banks are the wavelet coefficients at a desired level of resolution; for

Three-level wavelet analysis division of the time-frequency plane. Each rectangle represents one wavelet coefficient. The darker the rectangle's shading, the larger the magnitude of the wavelet coefficient.

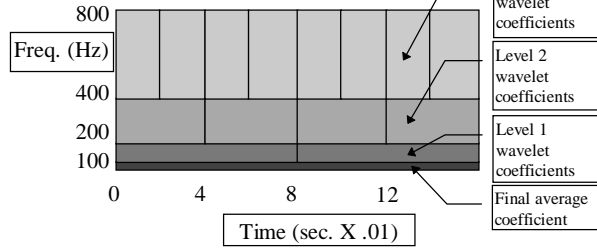


Figure 1

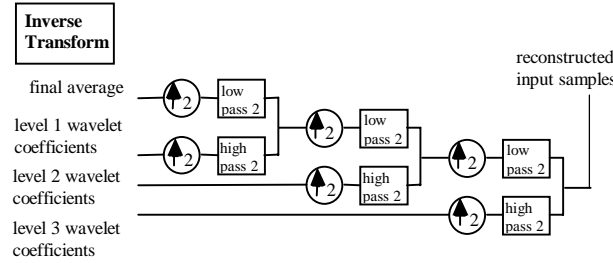
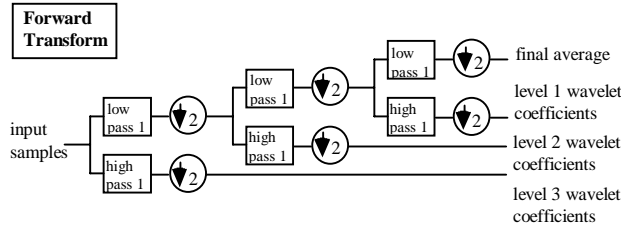


Figure 2

the inverse wavelet transform, the wavelet coefficients are the input, and the output of the final set of filter banks is the reconstructed window:

The construction of the four digital filters in Figure 2 is not a trivial task, as the filters must satisfy the dilation equation and several other mathematical restrictions, including non-distortion and non-aliasing requirements. In particular, developments in wavelet theory have established that requiring the low-pass filters in Figure 2 to have p zeros at π allows a filter to perfectly analyze and resynthesize a polynomial of degree $p-1$. Specifically, a filter that has p zeros at π is said to have p vanishing moments. Therefore, loosely speaking, the more vanishing moments a filter has, the more "jagged" a signal it can analyze and perfectly reconstruct.

There are many wavelet families whose analyzing and resynthesizing filter coefficients have already been derived, such as the Haar and Daubechies filters. However, in the present context, we require a filter

whose coefficients are symmetric with respect to a central axis, as we wish for no phase distortion to result from the filtering. Therefore, the wavelets and filters used in the excitation algorithm come from a family known as "binlets," which are biorthogonal, symmetric, binary coefficient filters. These wavelets and filters were chosen for their symmetry and simplicity [14].

3 Excitation algorithm

Choose a set of binlet analysis and resynthesis filters having p vanishing moments, where the audio signal to be excited is assumed to be roughly piecewise polynomial of degree $p-1$, and therefore smooth in this sense. Using the selected binlet analysis / resynthesis filters, perform an n th-level, dyadic, forward wavelet transform of the audio signal. Next, add a new level of wavelet coefficients to the existing wavelet analysis, one level above the most detailed level of the existing analysis. This will effectively double the frequency range and the amount of data in the analysis by creating an $(n+1)$ th level of wavelet coefficients.

We now note that Strang [14] states that the magnitude of wavelet coefficients decays exponentially across analysis levels, and that the rate of decay is related to the smoothness of the analyzed function and the wavelet analysis level:

- (1) If $f(t)$ has p derivatives, its wavelet coefficients at level j decay like 2^{jp} , where j is the wavelet analysis level, and

$$|b_{jk}| = \left| \int f(t) w_{jk}(t) dt \right| \leq C 2^{-jp} \|f^{(p)}(t)\|, \text{ where}$$

b_{jk} is the wavelet coefficient at level j , translation k ,
 w_{jk} is the wavelet basis function associated with level j , translation k , and
 $C \in \mathfrak{R}$

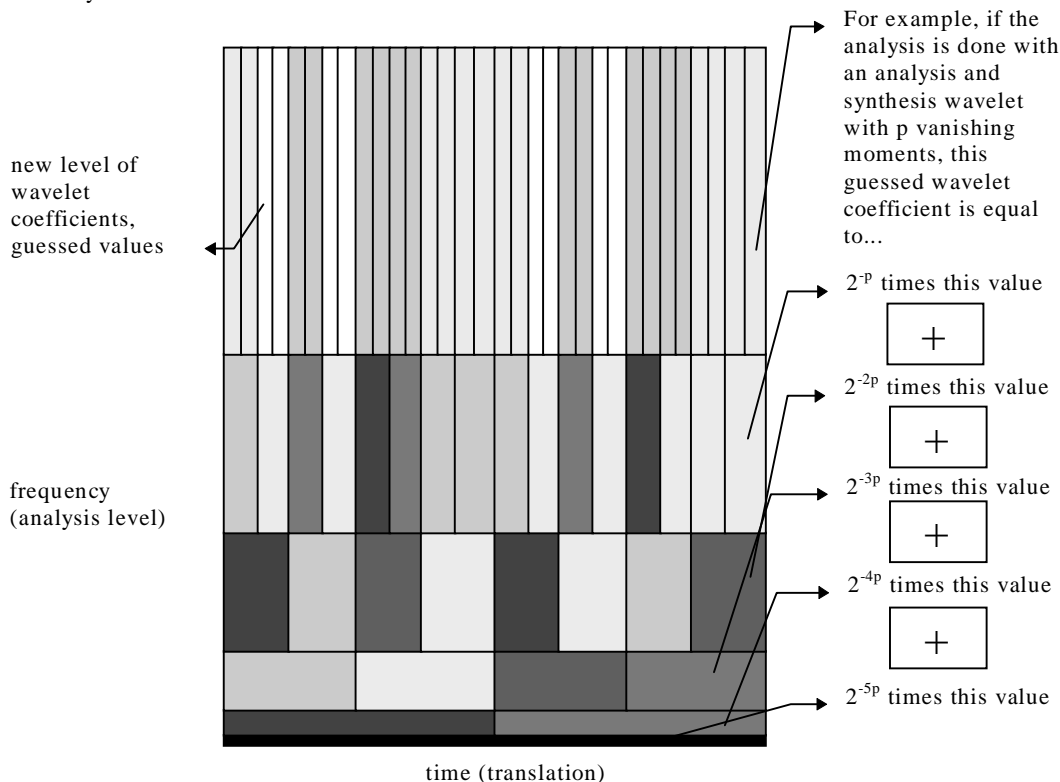
Consequently, extrapolate the $(n+1)$ th level wavelet coefficients' values from existing wavelet coefficients by assuming that the new coefficients are exponentially related to the previous levels' values. For example, in a five-level analysis, the guessed coefficient is a sixth-level coefficient which is equal to:

- (2) $b_6 = 2^{-p} b_5 + 2^{-2p} b_4 + 2^{-3p} b_3 + 2^{-4p} b_2 + 2^{-5p} b_1 + 2^{-6p} b_0$, where b_n is the "appropriate" wavelet coefficient at level n , and p is the number of vanishing moments in the analyzing and synthesizing wavelets

Equation (2) can be generalized into a formula which allows the guessed wavelet coefficient values to rely on fewer levels of wavelet coefficients than are present in the original analysis, to allow for some flexibility [4]:

Adding another level of wavelet coefficients to an existing analysis to increase the frequency bandwidth of a low-sample-rate audio file

Coefficient information is doubled, and the sample rate of the resulting file is also doubled to provide “headroom” for the new, higher frequency data. The new coefficients’ values are guessed from existing ones. A guessed coefficient value is determined from an exponentially decaying reliance on the coefficients that fall directly below it.



The number of levels upon which the guess depends can be chosen; typically, the more the dependence on lower levels, the more aliasing tends to occur. Most successful results were found when the guess only relied on one level: the level right beneath it.

Figure 3

(3)
$$b_{l+1,l} = \sum_{i=0}^N \frac{1}{2^{p(l-i+1)}} b_{l, \lfloor l/2^i \rfloor}$$
 , where

- $b_{s,t}$ is the wavelet coefficient at level s translation t
- N is the number of levels of wavelet coefficients desired e.g. $N = 1$ just relies on the previous level's coefficients
- p is the number of vanishing moments in the synthesizing and analyzing wavelets
- l is the translation of the guessed wavelet coefficient.
- l is the total number of iterations in the original wavelet decomposition
- $\lfloor \cdot \rfloor$ represents the least greatest integer (stair) function

Finally, perform an $(n+1)$ th level inverse wavelet transform on the $n+1$ levels of wavelet coefficients, normalize the result to account for power loss, and double the sampling rate associated with the data to account for the doubling of the amount of data.

4 Experiments and results

Original 44.1 kHz, stereo, 16-bit AIFF files were downsampled to 11.025 kHz in order to eliminate high frequency information in the process. The files were then spectrally enhanced twice to quadruple and restore the sampling rate to 44.1kHz. Qualitative comparisons with regard to sound quality were then made between the original file and the spectrally enhanced file. A window of 32768 frames was used for each example, and 2 iterations of the wavelet transform were performed for each lower resolution file. Best results were obtained when the guessed wavelet coefficients only relied on one level ($N=1$ in (3)). The 6/10 binary-coefficient wavelet analysis/ resynthesis filters were used ($p=3$ in (1)). Finally, the spectrally enhanced files were normalized to account for power loss, so that comparisons could be made to the original file.

The comparisons were good for drum sounds and high-pitched sounds, moderate for heavy percussion sounds, and relatively poor for speech sounds. This is consistent our expectations, since 1) speech is mostly contained in low-frequency regions, and therefore enhancement is essentially adding high frequency material that is not present in the original signal, and 2) wavelet analysis is well-suited to many types of transient analysis. Imaging was much improved in most cases, and ambiance was added in many cases as well. However, there was a grainy sound to the processed output, as the filter coefficients used did not have sharp rolloffs nor was any pre-processing done to the signal before performing the wavelet decomposition.

5 Conclusions and Future Directions

The results from this simple wavelet-based process are promising; frequency information has been added to low-sample-rate soundfiles that sound convincing in many cases. However, the algorithm presented here is only a first step which requires refinement in several ways: the power attenuation problem must be fixed; existing coefficients could also be altered to provide for a better result, and different, higher-order filters customized for audio analysis could be used.

6 Acknowledgments

The author would like to thank Professor Gregory H. Wakefield from the University of Michigan at Ann Arbor's Electrical Engineering Department for his help and encouragement in preparing this manuscript for publication. In addition, the author would like to thank the faculty and graduate students at the Bregman Electro-acoustic Music studios at Dartmouth College, where much of this research was done.

References

- [1] Arfib, D., and Delprat, N. "Musical Transformations Using the Modification of Time-Frequency Images." *Computer Music Journal*, 17:2. The MIT Press, Cambridge, MA: 1993.
- [2] Berger, J. and Nichols, C. "Using Wavelet Based Analysis and Resynthesis to Uncover the Past." *Proc. 1994 ICMC*, ICMA, San Francisco: 1994.
- [3] Chan, Y.T. *Wavelet Basics*. Kluwer Academic Publishers, Boston, MA: 1995.
- [4] Cheng, C. *Wavelet Signal Processing of Digital Audio with Applications in Electro-Acoustic Music*. Dartmouth College Masters' Thesis, Dartmouth College, Hanover, New Hampshire: 1996.
- [5] Chui, C. K. et al. *Wavelets: Theory, Algorithms, and Applications*. Academic Press, Inc., Boston, MA: 1994.
- [6] Cohen, A. and Daubechies, I. "Biorthogonal Bases of Compactly Supported Wavelets." *Communications on Pure and Applied Mathematics*, 45:485-560, 1992.
- [7] Daubechies, I. "Orthonormal Bases of Compactly Supported Wavelets." *Communications on Pure and Applied Mathematics*, 41:7. John Wiley & Sons, New York: 1988.
- [8] De Poli, G. et al., eds. *Representations of Musical Signals*. The MIT Press, Cambridge, Massachusetts: 1991.
- [9] Guillemain, Ph. and Kronland-Martinet, R. "Additive resynthesis of sounds using continuous time-frequency analysis techniques." *Proc. 1992 ICMC*, ICMA, San Francisco: 1992.
- [10] Kronland-Martinet, R. "The Wavelet Transform for Analysis, Synthesis, and Processing of Speech and Music Sounds." *Computer Music Journal*, 12:4. The MIT Press, Cambridge, MA: 1988.
- [11] Kussmaul, C. "Applications of the Wavelet Transform at the Level of Pitch Contour." *Proc. 1991 ICMC*, ICMA, San Francisco: 1991.
- [12] Kussmaul, C. *Applications of Wavelets in Music: The Wavelet Function Library*. Dartmouth College Masters' Thesis, Dartmouth College, Hanover, New Hampshire: 1991.
- [13] Popovic, I. et al. "Aspects of Pitch-Tracking and Timbre Separation: Feature Detection in Digital Audio Using Adapted Local Trigonometric Bases and Wavelet Packets." *Proc. 1995 ICMC*, ICMA, San Francisco: 1995.
- [14] Strang, G. and Nguyen, T. *Wavelets and Filter Banks*. Wellesley-Cambridge Press, Wellesley, MA: 1996.