

# Frequency Shaping of Audio Signals

Christopher Penrose  
Keio University SFC  
Faculty of Environmental Information  
5322 Endō  
Fujisawa, Kanagawa  
252-0816 Japan  
penrose@sfc.keio.ac.jp

## Abstract

*Frequency Shaping is a new high-resolution audio signal processing technique which effectively allows the frequency content of an audio signal to be tuned to the polyphonic frequency content of an entirely different audio signal. A result signal is produced by frequency shaping an amplitude reference and a frequency reference signal. The spectral formant of the amplitude reference signal will be accurately preserved in the result signal. However, the result signal will have the frequency content, or tuning, of the frequency reference signal. Frequency Shaping is non-heuristic and does not rely on classic pitch-detection or Linear Predictive Coding techniques. Frequency Shaping is also an efficient process which has been published in real-time software implementations (FFTease for Max/MSP) and used extensively in live-performance. Frequency Shaping is implemented utilizing a Short-Time Fourier Analysis and Resynthesis framework through use of a Fast Fourier Transform (FFT) algorithm. This paper will discuss the mathematics, implementation, and applications of this promising new signal processing technique.*

## 1 Overview

Linear Predictive Coding (LPC) has a long history of providing interesting and powerful audio processing (Moore 1988). Among other applications, the LPC analysis/resynthesis framework has been employed to create a variety of cross syn-

thesis and spectral warping techniques for pioneer computer musicians. LPC has been used in conjunction with monophonic pitch detection algorithms to map frequency information from a signal to the spectrum of an entirely different signal – the goal perhaps is to extend the possibilities and expressiveness of particular timbres. It has been used to great effect by many composers.

The premise of Linear Predictive Coding is to separate a signal's spectral formant and frequency content by representing a signal as a synthetic excitation of a complex infinite impulse response filter. The practical use of LPC processing in musical applications generally entails coupling linear predictive analysis with pitch detection – the pitch detection algorithm typically being monophonic. Monophonic pitch detection algorithms, though considerably more robust than those of the polyphonic variety, tend to be lacking in frequency resolution given the current state of the art.

The author has had a long interest in these particular applications of Linear Predictive Coding; he has often mused about polyphonic pitch remapping, polyphonic spectral formant warping and the possibilities they would present to digital music makers. While LPC techniques could potentially be coupled with high-resolution polyphonic pitch detection algorithms, the author has discovered a simpler, and perhaps more elegant technique which brings these polyphonic spectral manipulation ideas to fruition.

## 2 Analysis Framework

The Frequency Shaping process and its derivative tools are implemented by utilizing an overlapped Short-Time Fourier Transform (STFT) and its inverse. This framework is well documented in digital signal processing literature and further details can be found in the bibliography. See (Allen 1977 and Moore 1988).

For  $k = 0, 1, \dots, N/2 - 1$ , where  $\omega = 2\pi k/N$

$$F(k) = \sum_{n=0}^{N-1} w(n)I(n)e^{-i\omega n} \quad (1)$$

We perform a windowed Discrete Fourier Transform (DFT), as described by equation (1) above, upon overlapped segments of our real input signal  $I()$ . After representing our input segment in the frequency domain as a series of complex values, we convert these values into polar form:

$$M(k) = F_{mag}(k) = \sqrt{F_{real}(k)^2 + F_{imag}(k)^2} \quad (2)$$

$$P(k) = F_{phase}(k) = \arctan \frac{F_{imag}(k)}{F_{real}(k)} \quad (3)$$

We utilize a four-quadrant arctangent function, where the range of  $P(k)$  is  $-\pi, \dots, +\pi$ , to determine our phase angle,  $P(k)$ .

Utilizing polar form allows us to almost, but not quite, isolate signal frequency relationships from their magnitudes. It is important also to remember that spectral measurements are convolved with the transfer function of the analysis window,  $w(n)$ , effectively distributing frequency relationships of the analysis signal across multiple Fourier Analysis filters. Thus frequency measurements cannot be thought of as isolated to a single Discrete Fourier Transform analysis filter, but must be considered over a larger distribution of filter measurements. This particular phenomenon of windowing is pivotal to the success of Frequency Shaping.

The phase unwrapping technique of the phase vocoder is not necessary for our analysis/resynthesis implementation as phase relationships will not be altered from the successive analysis frames of the frequency reference signal. The Frequency Shaping process will be discussed at this stage of

STFT analysis, where our input signals are decomposed into a polar spectral representation, of the STFT analysis process.

We will analyze two separate time-domain input signals  $I_1(n)$ , and  $I_2(n)$  and generate an output time-domain signal  $O(n)$ , through a canonical inversion of the Short-Time Fourier Transform process.

## 3 Windowing

Frequency Shaping works by taking advantage of the convolution artifacts imparted by Discrete Fourier Transform (DFT) windowing. The frequency response of a sinusoid as measured by the DFT is a complex smear of frequencies – the spectrum measured by the DFT will be the convolution of the spectrum of the sinusoid with the spectrum of the analysis window,  $w(n)$ , along the frequency axis. The nature of this convolution is of particular interest.

The frequency response of a windowing function can be determined by calculating its  $z$ -transform. The DFT has an implicit window function, the rectangular window, which has the following frequency response  $W(\theta)$ :

$$W(\theta) = e^{-i(N-1)\theta/2} \frac{\sin(N\theta/2)}{\sin(\theta/2)} \quad (4)$$

As is described extensively throughout digital signal processing literature, modifying this window function significantly alters the frequency response of a DFT analysis of input signal  $I(n)$ . Frequency Shaping has been implemented and widely tested utilizing a Hanning window, where  $w(n)$  is constructed using the following:

For  $n = 0, 1, 2, 3, 4, 5, \dots, N - 1$

$$w(n) = .5 + \frac{\cos(2\pi n/N)}{2} \quad (5)$$

$W(\theta)$  for the above Hanning function is easily derived from equations (4) and (5) above. See (Steiglitz 1996) for more details.  $W(\theta)$  will scatter the the localized spectral measurement of  $I(n)$  over the entire range of DFT filters in the set  $F(k)$  by a relative of equation (4). The choice of  $w(n)$  will dictate the width of the central lobe of  $W(\theta)$ , the portion of the transfer function that will be centered at each approximately sinusoidal peak of the input,  $I(n)$ , which has the greatest local spectral energy.

## 4 Frequency Shaping

The goal of Frequency Shaping is to preserve the frequency distribution of the frequency reference signal,  $F_f(k)$ , while reflecting the spectral formant of the amplitude reference signal,  $F_a(k)$ . Even for complex enharmonic input, the implicit convolution of the window frequency response with the spectrum of  $I_f(n)$ , as a result of Fourier Transform processing, will allow frequency information from the analysis magnitude spectrum,  $M_f(k)$ , to be extracted or preserved.

Our process will focus upon a result spectrum,  $F_R(k)$ . We establish the range of DFT filters,  $w$ , that is used to determine an average ratio between the frequency reference magnitudes,  $M_f(k)$ , and the amplitude reference magnitudes,  $M_a(k)$ . This ratio is then used to scale  $M_f(k)$  creating the new magnitude spectrum  $M_R(k)$ . The effect of scaling  $M_f(k)$  by this ratio is to extract the spectral envelope from  $M_a(k)$  and impart it upon  $M_f(k)$  while maintaining the localized frequency information inherent to  $M_f(k)$ . The presence of  $M_f(k)$  in the denominator of equation (6) ensures that  $M_R(k)$  will have the average magnitude of  $M_a(k)$  rather than  $M_f(k)$  over  $w$  DFT filters.

The polar phases of  $P_R(k)$  are set to the polar phases of the frequency reference spectrum,  $P_f(k)$ . The result spectrum,  $F_R(k)$ , is created by transforming  $M_R(k)$  and  $P_R(k)$  into complex form, and is then resynthesized through use of an Inverse Discrete Fourier Transform (IDFT). Equations (6), (7) and (8) describe the Frequency Shaping process.

Our function,  $S(j)$ , encapsulates the ratio of  $M_a(k)/M_f(k)$  over  $j$ .  $k'$  is an integer index of  $S(j)$ .

For  $j = 0, 1, 2, 3, 4, 5, \dots, \frac{N/2-1}{w}$

$$S(j) = \frac{\sum_{n=0}^w M_a(jw + n)}{\sum_{n=0}^w M_f(jw + n)} \quad (6)$$

For  $k = 0, 1, 2, 3, 4, 5, \dots, N/2 - 1$ , where  $k' = k/w$

$$M_R(k) = M_f(k)S(k') \quad (7)$$

$$P_R(k) = P_f(k) \quad (8)$$

As the frequency content of the frequency reference signal,  $F_f(k)$ , is adequately represented by its localized amplitude distribution (the magnitude distribution within the range of  $w$ ) and its phases,  $P_f(k)$ ,  $F_R(k)$  will now have the frequency content of  $F_f(k)$ . This process succeeds for a broad

range of complex polyphonic signals because the width of the central lobe of our window function spectrum,  $W(\theta)$ , is broad enough to accommodate frequency localization, and its subsequent drop-off is extreme enough such that filters outside of the central lobe will not significantly contribute to frequency localization.  $w$  should be set to a value that approximates the size of the central lobe of  $W(\theta)$ . Keeping  $w$  as small as practical ensures that  $F_R(k)$  will reflect the spectral formant of  $F_a(k)$ . As  $w$  approaches  $N/2$ ,  $F_R(k)$  will instead reflect more of the spectral envelope of  $F_f(k)$ . Using a Hanning window, equation (5), practical values for  $w$  are in the range of 3-5 filters.

Our choice of DFT size,  $N$ , is also an important consideration. As larger values of  $N$  localize spectra with greater frequency resolution (at the expense of temporal resolution), smaller sizes of  $N$  tend to provide better frequency shaping results. If frequency resolution is too localized,  $M_f(k)$  and  $M_a(k)$  may have significant spectral differences preventing an effective exchange of frequency information. Also consider the ramifications of equation (4): as  $N$  increases, the size of the central lobe of  $W(\theta)$  decreases with respect to  $N$ , as the central lobe size remains near constant for all values of  $N$ . Thus, there is a practical limit to adjusting  $w$  as  $N$  increases, further diluting the ability of the process to be effective with large sizes of  $N$ . For  $R = 44100$ , a DFT size of  $N = 256$  is usually most effective.

## 5 Spectral Warping

Frequency Shaping has also been used to great success to create other useful forms of audio signal processing. Spectral Warping, the stretching or contraction of the spectral envelope of a signal, can be easily and accurately accomplished by use of Frequency Shaping with far greater clarity and transparency than through the use of Linear Predictive Coding. The polyphonic frequency distribution of the input signal is effectively maintained in the output signal.

A single input signal,  $I(n)$ , is analyzed utilizing the STFT. Utilizing a sample rate conversion algorithm, the polar magnitudes of our analysis,  $M(k)$ , are then companded (expanded or contracted), according to a scalar warp factor,  $warp$ , creating a modified spectrum  $M_{warp}(k)$ . The original signal magnitudes,  $M(k)$ , are then used to frequency shape the warped spectrum,  $M_{warp}(k)$ , after deriving it from sample rate conversion. For contraction cases, a filter is applied to spectral re-

gions in the original spectrum that no longer have energy: the magnitudes of energyfree spectral bands are set to an epsilon value. Those energyful bands close to the transition receive a shelving interpolation filter to ease the spectrum's approach to this epsilon value. At this point,  $M_{warp}(k)$ , will reflect the frequency content of  $M(k)$ , but its spectral envelope will have been companded by the warp factor,  $warp$ .

## 6 Time-Varying Filtering

Frequency Shaping has also been utilized to improve Finite Impulse Response filtering processors which utilize the Discrete Fourier Transform. Arbitrary functions can be applied to Fourier Transform magnitude data,  $M(k)$ , and to prevent frequency distortion, the filtered magnitudes can be frequency shaped with the original magnitude spectra. Though Frequency Shaping will decrease the resolution available for filtering by a factor dependent upon its averaging width,  $w$ , many Fourier Transform based FIR filtering applications do not suffer as readily from using larger transform sizes. And real-time filtering applications can be implemented with acceptable frequency resolution (1024 – 4096 FFT sizes) allowing use of arbitrary multi-band time varying filters given the adaptive frequency protection of this technique. The DFT size restrictions that effect the frequency remapping application of Frequency Shaping do not apply.

## 7 Software Implementations

Frequency Shaping is implemented as a UNIX commandline processor, *shapee*, which is found in the audio signal processing distribution, *PVNation*, and it is also implemented as a Max/MSP object, *shapee~*, found in the FFTease spectral processor collection. Information regarding this software can be found at via the World Wide Web:

<http://www.sfc.keio.ac.jp/~penrose/>

Frequency Shaping examples can also be found via the web:

<http://www.sfc.keio.ac.jp/~penrose/shapee/>

## 8 References

1. Jont B. Allen, "Short Term Spectral Analysis, Synthesis, and Modification by Discrete Fourier

- Transform." IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP-25(3), pp. 235-238 (1977)
2. Brown, J. C. and Puckette, M. "A High-Resolution Fundamental Frequency Determination Based on Phase Changes of the Fourier Transform." Journal of the Acoustical Society of America 94/2, pp. 662-667. (1993)
3. R. E. Crochiere, "A Weighted Overlap-Add Method of Short-Time Fourier Analysis/Synthesis", IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP-28(1), pp. 99-102 (1980)
4. J.-K. Hang and Y.-C. Chen, "Superresolution frequency estimation by alternating notch periodogram", IEEE Trans. Signal Proc., feb, vol. 41, No. 2, pp. 727-741 (1993)
5. F. Richard Moore, "Elements of Computer Music." Prentice-Hall, (1988)
6. James A. Moorer, "The Use of the Phase Vocoder in Computer Music Applications." Journal of the Audio Engineering Society 24(9). pp. 717-727 (1978)
7. Christopher Penrose, "Practical Signal Processing: Filtering, Interpolating, and Enriching Digital Signals with the Handy Phase Vocoder Algorithm." Proceedings of the Intl. Computer Music Conference and Festival at Delphi, Greece, CCMR Athens, Greece (1992)
8. Christopher Penrose, "Extending Musical Mixing: Adaptive Composite Signal Processing." Proceedings of the International Computer Music Conference (ICMC), Beijing, China (1999)
9. Michael R. Portnoff, "Time-Frequency Representation of Digital Signals and Systems Based on Short-Time Fourier Analysis." IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP-28(1), pp. 55-69 (1980)
10. Ken Steiglitz, "A Digital Signal Processing Primer." Addison-Wesley (1996)