Proceedings of the 2nd COST G-6 Workshop on Digital Audio Effects (DAFx99), NTNU, Trondheim, December 9-11, 1999

MUSICAL APPLICATIONS OF DECOMPOSITION WITH GLOBAL SUPPORT

Øyvind Hammer, Henrik Sundt

NoTAM
University of Oslo
P.O.Box 1137 Blindern, N-0317 Oslo, Norway
oyvindha@notam.uio.no

ABSTRACT

Much of today's musical signal processing is based upon local decomposition methods, often using short-range, windowed Fourier transforms. Such methods are supposed to mimick some aspects of human hearing. Using global decomposition instead, analyzing the complete source sound in one transform, opens up a new world of interesting sound manipulation methods. These methods are difficult to analyze in terms of human auditive perception or musical intuition but can produce exciting sounds. A program that explores these possibilities has been written and placed in the public domain.

1. INTRODUCTION

Signal decomposition using a set of orthogonal basis functions is fundamental to much of today's musical signal processing. The Fourier basis is the most familiar, but more general wavelet decompositions have entered the scene in recent years. The rationale for such decomposition is that sound processing in the transformed domain may be easier and involve parameters that are more powerful. For example, pitch shifting is a procedure that it seems natural to perform in the frequency domain. Usually, the basis functions have *local support*, meaning that they are non-zero only over a limited duration window. Windowed DFT (Discrete Fourier Transform) is supposed to mimick the way the ear processes sound, presenting short-term variation in air pressure in the frequency domain, while long-term dynamics are presented as a progression in time of spectra from consecutive windows.

For musical applications, "standard" operations such as pitch shifting, time stretching, filtering, spectrum shifting and cross synthesis are obviously needed, and decomposition with local support is a useful method for accomplishing them. Still, a plethora of new possibilities opens up if we let the basis functions have *global* support. This is a relatively non-intuitive approach, as it does not mimick any aspect of the processing of sound by ear and brain, but the auditory results can be unexpected and often very interesting.

2. THE LONG-TERM DISCRETE FOURIER TRANSFORM (LTDFT)

The idea that a DFT can be taken over the complete sound, instead of separately and successively in a number of small windows, was used by the Swedish composer Paul Pignon in his piece "Z", produced at EMS in Stockholm in 1987. He called this technique "Mammoth FFT", and implemented it on a VAX 11/750 with an array coprocessor. In “Z”, the spectral coefficients themselves were used as sonic material.

The concept of such global decomposition may seem strange. The complete sound is subjected to one single DFT, producing a number of complex spectral coefficients that is equal to half the number of samples in the sound. In this single spectrum, all time and frequency information in the sound is embedded. Dynamics that we perceive in the time domain are encoded in the phase and amplitude relationships between close frequency components, beating in complex patterns to produce the development in time. Any alterations to the phases or amplitudes of the partials may have large, non-intuitive consequences in the time domain.

The DFT is conveniently computed using the Fast Fourier Transform. In radix-2 FFT, the signal is supposed to have a number of samples that is equal to some power of two. The signal is therefore zero-padded to contain a correct number of samples. Since the FFT has a computation time that is proportional to $N \log N$, where N is the number of samples, total computation time for a windowed FFT of a given sound increases as $\log W$, where W is the window size. This means that an FFT with global support is more computationally expensive than splitting the signal into windows and decomposing these separately.

3. SOUND ALTERATION WITH THE LTDFT

Many different sound processing methods have been tested in the
LTDFT frequency domain. Generally, it seems that the most interesting transformations involve changes in phase and/or non-linear adjustment of the positions of partials. Some particular transformations are listed below.

Stretch. All frequencies (FFT bin numbers) are raised to the power of the exponent specified, and the frequency axis is then re-normalized. When the spectral coefficients are mapped to other FFT bins, no interpolation is carried out, leading to loss of some coefficients in compressed areas and introduction of zero bins in stretched areas. This non-linear stretching of the frequency axis produces dispersion effects with frequency sweeps. An informal argument shows that the effect is due to the fact that a compression or expansion of the frequency axis corresponds to a reciprocal change in the time axis:

$$x(at) \leftrightarrow X(\omega / a) / |a|. \tag{1}$$

By raising the frequencies to the power of a positive exponent, the frequency axis is progressively more stretched at higher frequencies (increasing a), corresponding to a larger compression of the time axis. In this way, higher frequency components occur earlier than the lower components, producing descending sweeps. Conversely, using a negative exponent produces ascending sweeps.

Wobble. This transform will alternately stretch and contract the frequency axis using a sinusoidal transfer function for the frequencies. The *Frequency* parameter controls the number of periods of the transfer function from 0 Hz to the Nyquist frequency, while *Amplitude* controls its amplitude (1 is the entire frequency axis). This causes a frequency-dependent time scaling with both ascending and descending frequency sweeps.

Multiply phase. Multiplies all phases with the value specified. A value of -1 will reverse the sound. This simple transform can produce highly surprising and complex results. Multiplying all phases by a constant k implies that a unit impulse $\delta(t-T)$ at time T is moved to a new position kT, wrapping around if kT is greater than the analysis length T_A. Calling the phase multiplication operator Φ_k, we have

$$\Phi_k[\delta(t-T)] = \delta(t-(kT \bmod T_A)). \tag{2}$$

Since the delay is dependent upon the original position of the impulse, this is not a time-invariant operation. Parts of the original sound are moved around and superimposed onto eachother, keeping continuity at all times and thus avoiding clicks or noise.

Derivate amp. Replaces the amplitude spectrum with its derivative (slope). This will highlight rapid changes along the frequency axis. The result is a time development similar to the original sound, but with a highly reverberant background.

Filter. 'Maximally sharp' bandstop filter, often giving substantial ringing.

Invert blocks. Splits the spectrum into regions with specified size, and turns each of these backwards. If a region size of 100% is selected, the entire spectrum will be mirrored around its center. Large parameter values (1-100%) produce effects which are perceived in the frequency domain, while small values (<1%) produces effects perceived as a multitude of delays and reversals.

Threshold. Removes all partials below a given amplitude threshold. This rarely gives effects perceived in the time domain.

Spectrum shift. Simple spectrum shift, with no window artefacts. The frequency specified (positive or negative) will be added to all frequency values.

Block swap. Selects randomly positioned regions of the spectrum, and interchanges their halves. The *Block size* parameter sets the size of the blocks, given in percents of the whole frequency axis. This procedure is repeated a number of times, as specified, thus permutating the partials. Depending on the parameters, this can produce effects that are perceived either in the frequency or the time domain, or both.

Mirror. Reflects the whole spectrum around the frequency specified. This does not produce any effect in the perceived frequency domain.

4. CROSS SYNTHESIS

Cross synthesis can also be carried out in the LTDFT frequency domain, by multiplying two spectra in different ways. Standard, high-quality convolution with e.g. a room response involves complex multiplication of the two spectra. Correlation is implemented as a similar complex multiplication but with one spectrum complex-conjugated. Other operators can also be implemented.

5. THE "MAMMUT" PROGRAM

A program with a graphical user interface has been developed to implement the algorithms described above. "Mammut" (Norwegian spelling of mammoth) is available as freeware for SGI and Linux. The program displays the LTDFT spectrum, and allows processing of the spectrum using menus and dialogue boxes. Mammut has been used in the production of a number of computer music pieces at NoTAM.

The latest version of Mammut for SGI is available at `http://notam.uio.no/notam/mammut-e.html`.

The Linux version (ported by Dave Phillips) can be found at `ftp://mustec.bgsu.edu/pub/linux/mammut-linux.tar.gz`.

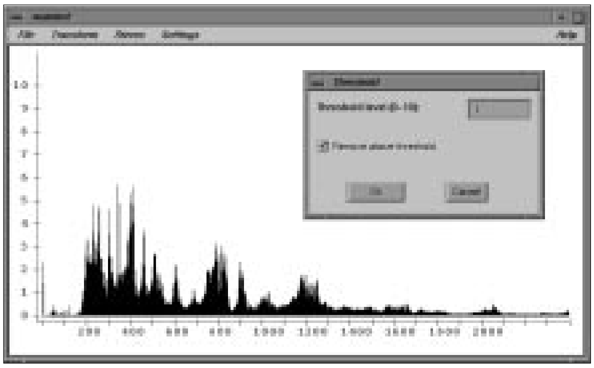

Figure 1. The Mammut program.