Extending the McAulay-Quatieri Analysis for Synthesis with a Limited Number of Oscillators

Kelly Fitz, William Walker, Lippold Haken

CERL Sound Group, University Of Illinois in Urbana-Champaign (k-fitz@uiuc.edu)

Abstract

The McAulay-Quatieri (MQ) analysis is a robust, general sinusoidal analysis technique. Unlike many other analysis techniques, it can be used to analyze sounds without a stable harmonic structure (i. e., polyphonic or non-harmonic sounds and instrument tones with extreme vibrato). The MQ technique can provide the time-varying spectral information needed to control a real-time additive synthesis engine. Unfortunately, the original MQ technique generates an arbitrary number of sinusoidal tracks, while a real-time system has a limited number of oscillators. This paper presents some improvements to the basic MQ analysis technique which, in addition to improving the quality of the ensuing syntheses, and making the sinusoidal model more robust, make the MQ analysis data more suitable for controlling a real-time sinusoidal synthesis engine with a fixed number of oscillators.

1. The Choice of the McAulay-Quatieri Model

In our research, we have sought a model for sound that would accommodate synthesis with time scale modifications. The goal has been to find a model that would produce time-scaled syntheses of sampled audio signals that are otherwise perceptually equivalent to the original signals. We chose a sinusoidal model advanced by McAulay and Quatieri (McAulay and Quatieri 1985) as the basis for our research. The MQ sinusoidal model allows independent time- and frequency- scale modification.

Many implementations of sinusoidal modeling derive sinusoidal components from the Short-Time Fourier Transform (STFT) and are of limited use for time- and frequency-scale modification because of artifacts caused by phase uncertainty and discotinuities. (A more detailed description of the STFT may be found in signal processing texts.) Pitch tracking analysis methods have been used in the past to solve some of these problems, but their use is restricted to the class of monophonic, strongly harmonic signals (Grey 1975, Haken 1989).

McAulay and Quatieri (McAulay and Quatieri 1985) propose a sinusoidal analysis technique for speech processing. The premise of the MQ technique is that a sound can be represented by a collection of sinusoidal components (called tracks), each with time-varying amplitude and frequency. To construct these tracks, STFT's are performed on a signal at regular intervals, called frames. Amplitude peaks in the resulting frequency spectra are identified, and parabolic interpolation is used to obtain a close approximation of the exact spectral peak frequencies. These peaks are the most prominent frequencies in the sound at that instant. The peaks in adjacent frames are compared and peaks of similar frequencies are matched. A continuous chain of these matched peaks is a track. A peak that is not matched represents the birth or death of a track. MQ synthesis uses cubic phase interpolation to reduce phase uncertainty and eliminate phase discontinuities.

2. Lemur Extensions

Lemur is a Macintosh implementation of the MQ technique with some extensions. It is based on the program MQAN, written by Rob Maher and James Beauchamp at the Computer Music Project at the University of Illinois (Maher 1989), which implemented the original MQ analysis/synthesis technique on a UNIX system. Lemur provides some extensions to the basic MQ technique.

2.1 Frequency Bins

The original MQAN algorithm models psychoacoustic masking effects by defining the spectral peak magnitude threshold in terms of the difference in magnitude between the largest peak in the frame and the peak under consideration. Using this relative threshold, very quiet or silent portions of the sound, for which the spectrum is virtually flat and very low in magnitude, will produce an overwhelmingly large number of peaks. When resynthesized, these peaks sound like low amplitude hiss or background noise. To avoid this problem, an additional, absolute lower threshold can be imposed, and the final threshold for a frame is the maximum of the relative threshold, computed relative to the largest peak in the frame, and the absolute lower threshold, which is static over the entire analysis (Maher 1988).

Unfortunately, this dual threshold scheme ignores the importance of frequency in masking effects. Peaks of very different frequencies rarely mask each other, so a very quiet high frequency sinusoid will be perceived even in the presence of a very loud low frequency sinusoid. Lemur provides a refinement to the original dual magnitude threshold scheme by breaking the frequency domain into logarithmically-sized bins. The loudest peak in each bin is determined, and a relative threshold for each bin is computed based on its loudest peak. This allows quiet peaks to be ignored in a bin containing loud peaks, while detecting quiet peaks in a bin without loud peaks. The absolute threshold is applied globally across the frequency spectrum. Thus, the peak magnitude threshold for a particular frequency bin in a frame is the maximum of the relative threshold for that bin and frame, and the absolute threshold for the analysis. This is not a psychoacoustically accurate model of the effect of frequency in masking, but is an approximation which presents a significant improvement over the original model. A psychoacoustically accurate model is computationally prohibitive, and may not yield perceptibly more accurate syntheses.

The use of frequency bins in the MQ analysis results in the representation of many more high frequency components and markedly better sounding syntheses. Figure 1 shows track diagrams for two analyses of the same sound, one using one frequency bin and another using eight frequency bins. The horizontal lines on the graphs represent the tracks stored in the analysis.

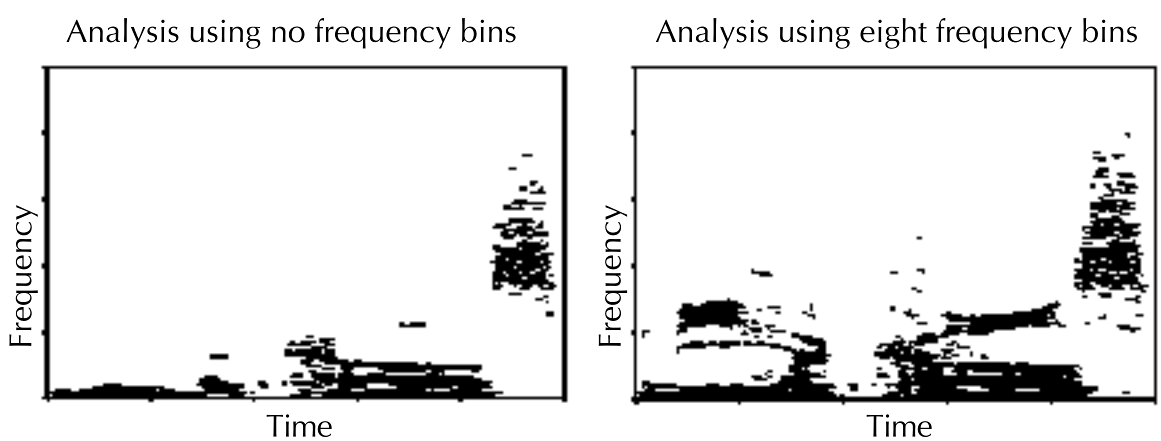

Figure 1

The use of frequency bins allows more significant tracks across the frequency spectrum to be included in the analysis data, without producing an unnecessarily large number of inaudible components. This is important when synthesizing with a limited number of oscillators.

2.2 Hysteresis

In examining the results of an MQ analysis, one often observes a track that dies out and another that is born a few frames later at roughly the same frequency. A series of such births and deaths at one frequency often indicates that several tracks are being used to represent a single sinusoidal component that is very close to, and periodically drops below the peak magnitude threshold. These are best understood as segments of the same track. Earlier attempts (Serra 1989) to facilitate this representation allowed tracks to lie dormant for a specified number of frames before dying out. A dormant track had zero magnitude, but still participated in track formation. The dormancy representation gave a more intuitive and visually-pleasing graph of the analysis, but did nothing to reduce the audible effects of low amplitude tracks repeatedly dying and being reborn, because peaks below the magnitude threshold continued to be synthesized at zero magnitude (this has been affectionately called the "doodley-doo" effect).

Lemur reduces the "doodley-doo" effect by allowing the specification of a track magnitude hysteresis. This is the amount by which a track may dip below the magnitude threshold while still participating in synthesis. A track may not be born at a magnitude below the peak magnitude threshold. It may, however, drop below that threshold over the course of the synthesis. Hysteresis may also be understood as the use of two different peak magnitude thresholds, one for births and another for deaths. Hysteresis differs from dormancy in that the tracks in the hysteresis range are synthesized at the magnitude reported from the frequency spectra, rather than at zero magnitude.

The audible effects of using hysteresis are less remarkable than the improvements obtained from the use of frequency bins. The effects are most apparent in sounds with long decays or reverb. Figure 2 shows track diagrams for two analyses of the same sound, one with no hysteresis, and one with 15 dB of hysteresis. Since hysteresis does not add tracks to the sinusoidal model, it can be used to improve the quality of a synthesis without the risk of demanding additional oscillators.

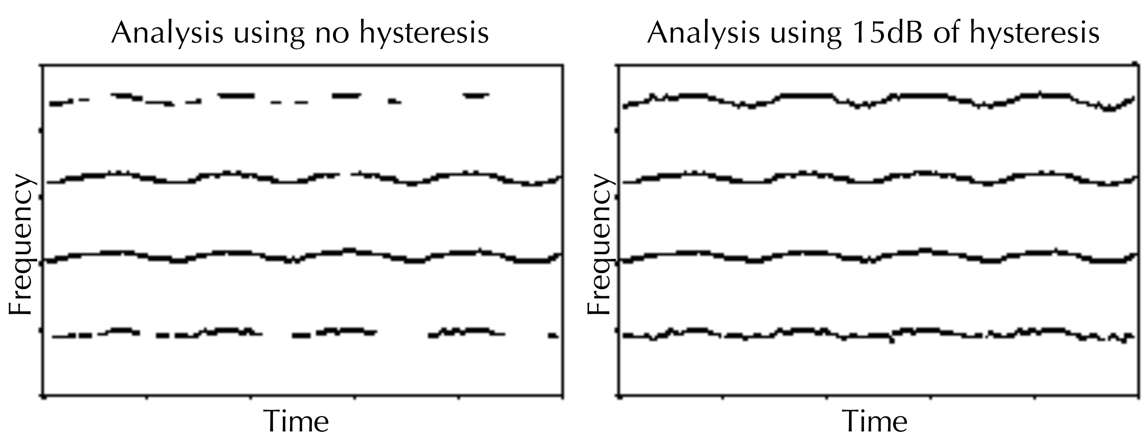

Figure 2

3. Conclusion

The McAulay-Quatieri technique for analysis and synthesis represents a robust sinusoidal model that is applicable to a broad class of sounds, and accommodates independent time- and frequency-scale modification. We have presented some improvements to the basic MQ technique that improve the quality of the synthesis and the intelligibility of the analysis data, and that make the technique suitable for real-time synthesis on a machine with a fixed number of sine wave oscillators.

4. Acknowledgments

This research was performed at the laboratory of the CERL Sound Group at the University of Illinois. The authors wish to acknowledge the work of Rob Maher and James Beauchamp at the University of Illinois Computer Music Project in developing the MQAN program, on which we based our research and the development of Lemur.

The figures for this paper were created using LemurEdit 1.0, written by Bryan Holloway at the CERL Sound Group, at the University of Illinois.

5. References

John Grey, *An Exploration of Musical Timbre.* Dept. of Music Report No. STAN-M-2, 1975, Stanford University.

Lippold Haken, *Real-time Fourier Synthesis of Ensembles with Timbral Interpolation.* Ph. D. dissertation, 1989, Dept. of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign.

Robert Crawford Maher, *An Approach for the Separation of Voices in Composite Musical Signals.* Ph. D. dissertation, 1989, Dept. of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign.

T. F. Quatieri and R. J. McAulay, *Speech Analysis/Synthesis Based on a Sinusoidal Representation.* Technical Report 693, Lincoln Laboratory, M. I. T., 1985

Xavier Serra, *A System for Sound Analysis/Transformation/Synthesis Based on a Deterministic Plus Stochastic Decomposition.* Dept. of Music Report No. STAN-M-58, Ph.D. dissertation, 1989, CCRMA, Stanford University.