

Research proposal: Extraction of relevant controllers for the “analysis by synthesis” of musical sounds

Sylvain Le Groux, Ph.D. student. MTG, Pompeu Fabra University

5th October 2004

Abstract

Psychoacoustics and the active field of Music Information Retrieval, are providing us with a lot of powerful techniques and insights for the analysis of musical sounds. Our proposal is to do some reverse engineering, and to find a practical way to control sound synthesis algorithms with intuitive high level parameters.

Outline

1. Motivation
2. Common psychoacoustics features
3. What we can learn from analysis/synthesis expertise
4. Tentative of general taxonomy
5. Ideas on mapping
6. Suggestions/Comments/Questions

Motivation

- Problem:
 - It is hard to really have a feel for what a lot of MIR features are describing. They can be statistically relevant for classification/recognition purposes but are often not really explanatory.
 - Help a non technically-oriented person to control sound synthesis.
- Proposal:
 - Usually the user feeds the synthesis engine with parameters. We want to control synthesis with higher-level descriptors directly related to the acoustic properties of sound.
 - Use the advents of MIR to provide intuitive synthesis controllers

- Why is it relevant to MIR field of research?
 - We can hear the influence of a variation of a feature on the synthesized sound. By doing so evaluate the perceptual relevance of this parameter. “I believe what I can hear”.
 - Possible “sketching of sound” for search queries in sound samples libraries
- Technology is here:
 - lot of well known synthesis techniques are efficient and available.
 - complex audio processing now feasible on a PC
 - machine learning tools widely available
- The missing links:
 - we need to define a corpus of perceptually relevant high-level controllers for synthesis

- we want the parameters to be automatically extracted if possible
- we need a method to map perceptual controllers to synthesis parameters

Psychoacoustics features

The parameters relevant to the listener are loudness, panning, pitch, timbre. Loudness, panning and pitch are pretty well known and defined.

- Loudness:
 - fechner law: sensation $\propto \log(\text{excitation})$
 - in fact **loudness** is a function of frequency too (Fletcher and Munson diagrams)
- Pitch:
 - 20 to 20000Hz. Frequency resolution is not constant. Depends on frequency. The **bark scale** representation is closer to our perception.
 - In fact it is even more complex as pitch depends also on loudness and duration of notes.

The main challenge of our project is to be able to control timbre, an ill-defined concept.

- Timbre:
 - by definition (or negation of a definition), timbre is what allows us to differentiate two sounds with the same loudness and pitch. The practice of analysis/synthesis of sound can help us extract important characteristics of timbre.

What we can learn from analysis/synthesis expertise

Practice and experience in analysis/synthesis of sound allows us to retain relevant information which has been aurally validated by the synthesis step.

- Timbre and spectrum, first approach
 - Any periodic waveshape is completely defined by amplitudes and phase of a harmonic series of frequency components
 - The effect of phase is usually inaudible. But this insensitivity only holds for phase relationship between harmonics of periodic tones.
 - In many cases, a fixed **formant** structure gives timbre that varies less with frequency than a fixed spectrum.
 - **Brightness** is related to the position of the spectral envelope along the frequency axis
 - **Presence** relates to strong components around 2000Hz

- Limitations
 - musical tones are usually not strictly periodic.
 - even from a completely distorted spectra (bad recording) a sound characteristic can be recognized
 - of course temporal aspect of timbre to be taken into consideration too
 - synthesis is a good means to control the aural relevance of the features extracted in the analysis
- Instrumental additive synthesis
 - often for a given intensity, spectrum has a formant structure. We want a roughly invariant **spectral envelope**. Piecewise amplitude envelope is enough.
 - **attack-time** is important
 - subjective **warmth** linked to the harmonicity of the partials
 - slow frequency modulation or **vibrato** enhance naturalness

- Percussive sounds synthesis
 - partial are no longer harmonics
 - realism increased by introducing slow **amplitude modulation** for certain components of the spectrum (bells)
 - adding time-varying **noise**
- Cross-synthesis
 - the experiments of exchanging spectral and temporal envelopes, indicated that they were at least as important.
 - sharp attack with exponential decay gives a **percussive quality** to any waveform. In this case, temporal envelope dominates over spectral one.
- FM
 - the amount of modulation (modulation index) is linked to the **width** of the produced spectrum

- Physical models
 - model is controlled through really intuitive physical parameters (length, tension, ...). The problem is that each type of sound requires a different model that is not easy to find and parametrize. Notion of **physical feature of the sound source**

General high level taxonomy

A possible definition of a general taxonomy for sounds has been proposed by P.Schaeffer. It is actually used in AudioClas project for analysis and retrieval purposes. This could be really useful for user-friendly synthesis too, since this taxonomy is supposed to help define a sound without getting too technical.

The taxonomy is based on the pair shape/matter. The matter criteria correspond to what we would hear if we could freeze the sound, while the shape is related to the variation of matter in time.

- Matter Criteria
 - mass
 - **pitch**ness of a sound
 - Harmonic timbre
 - **bright**ness

- Grain:
microstructure of the sound

- Shape criteria
 - Dynamics:
evolution of **loudness**
 - Allure:
Amplitude or frequency modulation.

- Variation criteria
 - Melodic profile:
Type of pitch variation
 - Mass profile:
Type of mass variation.

The descriptors we studied before (loudness, pitch, brightness, presence, spectral envelope, attack time, warmth, vibrato, amplitude modulation, noise) are directly related to Schaeffer's taxonomy.

Ideas on mapping

- Spline and spectral envelope
 - spectral envelopes can be approximated by splines
 - then, parametrized spectral envelopes can be controlled by spectral centroid (center of gravity), spectral spread(variance), skewness(symmetry). The model is not instrument/sound dependant.
- ANN
 - experiments on artificial neural network show that it is possible to control synthesis parameters from loudness, pitch and brightness. The learning is done on a specific set of data for one type of sound. Not sure if it can have a great generalization power...
- Genetic algorithms

- search space made of basic synthesis functions (filter, addition, sine, noise generator, ...). genetic algorithms find an appropriate combination for a given task.

Suggestions/Comments/Advices/Questions...