MUSICAL SOUND EFFECTS IN THE SAS MODEL

Sylvain Marchand

SCRIME - LaBRI - Université Bordeaux I
351 cours de la Libération, F-33405 Talence Cedex, France
sm@labri.u-bordeaux.fr

ABSTRACT

Spectral models provide general representations of sound in which many audio effects can be performed in a very natural and musically expressive way. Based on additive synthesis, these models control many sinusoidal oscillators via a huge number of model parameters which are only remotely related to musical parameters as perceived by a listener. The Structured Additive Synthesis (SAS) sound model has the flexibility of additive synthesis while addressing this problem. It consists of a complete abstraction of sounds according to only four parameters: amplitude, frequency, color, and warping. Since there is a close correspondence between the SAS model parameters and perception, the control of the audio effects gets simplified. Many effects thus become accessible not only to engineers, but also to musicians and composers. But some effects are impossible to achieve in the SAS model. In fact structuring the sound representation imposes limitations not only on the sounds that can be represented, but also on the effects that can be performed on these sounds. We demonstrate these relations between models and effects for a variety of models from temporal to SAS, going through well-known spectral models.

1. INTRODUCTION

In order to synthesize new digital sounds or manipulate existing ones using a computer, we need a formal representation for audio signals. A sound model constitutes such a mathematical representation. Although other types of models exist, this article focuses on spectral models, since they provide general representations of sound in which many audio effects can be performed in a very natural and musically expressive way. Based on additive synthesis, spectral models control many sinusoidal oscillators via a huge number of model parameters which are only remotely related to musical parameters as perceived by a listener.

The Structured Additive Synthesis (SAS for short) sound model [1, 2] keeps most of the flexibility of additive synthesis while addressing this problem. It consists of a complete abstraction of sounds according to only four physical parameters closely related to perception: amplitude, frequency, color, and warping.

Whereas the design and control of most of the audio effects becomes considerably simplified, some effects turn out to be impossible to achieve in this model. In fact structuring the sound representation imposes limitations not only on the sounds that can be represented, but also on the effects that can be performed on these sounds. There is a kind of trade-off of complexity versus feasibility in every sound model.

We demonstrate these relations between models and effects for a variety of spectral models in section 2, listing their main advantages and drawbacks and focusing on the feasibility and the complexity of sound effects in these models. Section 3 briefly presents the SAS model and the way sounds get represented in this model, while section 4 focuses on the design and control of musical sound effects in this model.

2. SOUND MODELS AND EFFECTS

It appears that structuring a model in order to facilitate the design of some kinds of sound transformations gives rise to both restrictions on the sounds that can be represented and impossibilities for other kinds of transformations.

2.1. Temporal Model

In the temporal model $a(t)$ represents the acoustic pressure at one point in space. The stream of samples of the discrete version of a can be sent directly to the digital to analog converter of any sound card. Another advantage of this representation it that everything is possible: there is no restriction on the sounds that can be reproduced or on the transformations that can be performed. Its main drawback is probably that none of these transformations are really musically intuitive. Of course amplification can be performed simply by multiplying the signal with an amplification factor, provided that it varies "slowly", or else amplitude modulation occurs, thus changing the sound timbre. But pitch transposition while preserving duration is a real challenge, and time-stretching without pitch-shifting is another one. The temporal model is however well-suited for effects like echo or reverberation, as they involve a superposition of pressure waves replicas in time. When the impulse response of the room is known, reverberation is similar to filtering. Filtering a sound requires either the composition of a filtering function with a, or convolving a with the impulse response of the filter, or even switching to the Fourier domain and multiplying their respective spectra.

2.2. Spectral Models

Spectral models attempt to parameterize sound at the human receptor, more precisely at the basilar membrane of the ear. Since spectral models are closer to perception, the design of sound effects should be more musically intuitive.

2.2.1. Phase Vocoder

The phase vocoder [3] uses the short-time Fourier transform to produce series of short-term spectra taken on successive windowed temporal frames. The problem is that there are many physical parameters that are not musically relevant such as the type, size and hop of the analysis window, and these parameters have a great impact on the analysis precision. The main advantage is that the sounds are now represented in a spectral domain. As a consequence, filtering gets trivial: its is just a matter of multiplication among spectra once the spectral response of the filter is designed, which is quite easy to do.

2.2.2. Additive Synthesis

The McAulay-Quatieri analysis, implemented in Lemur [4], looks across short-term spectra for partials. These partials are pseudo-sinusoidal tracks for which amplitudes and frequencies evolve slowly with time. The audio signal a can be calculated from the additive parameters using equations:

$$a(t) = \sum_{p=1}^{P} a_p(t) \, cos(\phi_p(t)) \tag{1}$$

$$\phi_p(t) = \phi_p(0) + 2\pi \int_0^t f_p(u) \, du \tag{2}$$

where P is the (finite) number of partials and f_p, a_p, and ϕ_p are respectively the instantaneous frequency, amplitude and phase of the p-ieth partial. The P pairs (f_p, a_p) are the parameters of the additive model and represent points in the frequency-amplitude space, as shown in Figure 1. Whereas time stretching gets trivial in this model, reverberation turns out to be impossible. In fact the reverberation of a partial is not a partial anymore. Indeed a single partial can lead to a huge – possibly infinite – number of simultaneous frequencies, so that the reverberated sound would not be in the model anymore (since the number of partials P would not be finite). Of course reverberation is a linear transformation, and does not produce frequencies that were not in the original sound. But this is not a "short-time linearity" in a sense that there can be at a certain time t frequencies that did not exist at time t in the original sound. Those frequencies are replicas of frequencies emitted before t.

2.2.3. Spectral Modeling Synthesis

Additive synthesis can faithfully reproduce a wide variety of sounds, but can not produce noises or transients. SMS [5] (*Spectral Modeling Synthesis*) adds noises to the additive model, while S+T+N [6] (*Sinusoids+Noise+Transients*) extends SMS with transients represented in the temporal model. These models are extremely expressive and allow perfect filtering or time stretching while simplifying the design of such effects too. However cross-synthesis or pitch-shifting without shifting formants require another level of structuration for the parameters.

3. THE SAS MODEL

The models based on additive synthesis are extremely difficult to use directly for creating and editing sounds. The reason for this difficulty is the huge number of model parameters which are only remotely related to musical parameters as perceived by a listener. The Structured Additive Synthesis (SAS) model keeps most of the flexibility of additive synthesis while addressing this problem. It imposes constraints on the additive parameters, giving birth to structured parameters closer to perception and musical terminology, thus reintroducing a perceptive and musical consistency back into the model. The remainder of this section quickly presents the SAS model. An extended presentation can be found in [1, 2].

3.1. Structured Parameters

SAS consists of a complete abstraction of sounds according to only four physical parameters, functions closely related to perception. These parameters – amplitude, frequency, color, and warping – are inspired by the work on timbre of researchers like Risset, Wessel, and McAdams, and by the vocabulary of composers of electro-acoustic music. We note (A, F, C, W) a sound in the SAS model. The first two parameters – amplitude A and frequency F – are unidimensional, functions of time only, while the two others – color C and warping W – are bidimensional, functions of both frequency and time. All these parameters vary slowly with time.

In the additive representation, the amplitude A corresponds to the sum of the amplitudes of all partials and for harmonic sounds F coincides with the fundamental, possibly missing or "virtual". Human beings perceive these parameters on a logarithmic scale. In such a scale, given a value of reference, amplitude is related to intensity which corresponds to the volume in dB while the frequency corresponds to the pitch. Color coincides with an interpolated version of the spectral envelope. We call it color after the analogy between audible and visible spectra already in use for noises (white, blue, etc.). Color and its manipulation is amply used in contemporary popular music, though such manipulations are inherently present in the timbre of some

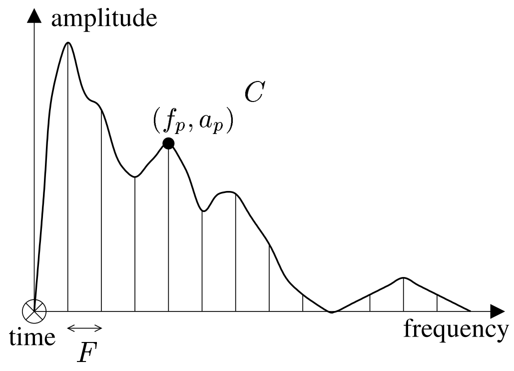

Figure 1: *An harmonic sound at time t.*

ancient instruments like the didjeridoo, which shows very unusual color variations. Harmonic sounds are totally defined by the A, F and C parameters (see Figure 1). But when sounds are not perfectly harmonic, the partial frequencies are not exactly multiples of the fundamental frequency F. That is why the fourth parameter – called warping after wavelet terminology – gives the real frequency of a partial from the theoretical one it should have had if the sound had been harmonic. Of course all harmonic sounds verify $\forall f, t, W(f,t) = f$.

To convert a sound from its temporal representation to the SAS model, one can perform a short-term Fourier analysis, then track partials across short-term spectra, to finally extract the SAS parameters from the set of partials, thus going through three levels of structuration. InSpect [7] can perform this conversion, using a new analysis method [8] in the first level of structuration to achieve sufficient precision.

3.2. Structured Equations

From the four structured parameters, we can calculate the the audio signal a in the temporal model using equations:

$$a(t) = A(t)\frac{\sum_{p=1}^{P} C(W(pF(t),t),t)cos(\phi_p(t))}{\sum_{p=1}^{P} C(W(pF(t),t),t)}$$

where $P = max_t\{\lfloor \frac{F_{max}}{F(t)} \rfloor\}$
(F_{max} is the highest audible frequency) and

$$\phi_p(t) = \phi_p(0) + 2\pi \int_0^t W(pF(u),u)\,du$$

These equations are the "structured" version of equations 1 and 2, and require approximately the same computation time. Any sound can be faithfully synthesized in real time from the model parameters using these equations. The underlying real-time additive synthesis has been implemented in the ReSpect software tool [7].

3.3. Model Restrictions

SAS can faithfully reproduce a wide variety of sounds – as additive synthesis does – provided they are monophonic sources. However it can not produce noises or transients. Noises can be added to SAS in the same way as in SMS, since every noise can be modeled as a filtered (colored) white noise at a certain amplitude. The amplitude and color parameters exist also for noises and are sufficient to define any noise. White noise has a white color ($C = 1$), and every noise named after an analogy with a light spectrum matches this correspondence of terminology.

4. MUSICAL SOUND EFFECTS

The presentation will show that in the SAS model many effects become accessible not only to engineers, but also to musicians and composers. Among these are filtering, time stretching, cross-synthesis, morphing, etc. These effects turn out to be straightforward in the SAS model and can be designed in a very intuitive way. This is done at the expense of restrictions not only on the kind of sounds that can be represented in the SAS model, but also on the kind of effects that can be performed.

4.1. Transforming Sounds

Given an SAS sound $S = (A, F, C, W)$, the simplest sound transformations can be expressed as a simple multiplication on one of its SAS parameters. In the remainder of this section we use the standard notation for the product between functions, that is for unidimensional parameters $(K \cdot P)(t) = K(t) \cdot P(t)$ while for bidimensional ones $(K \cdot P)(f,t) = K(f,t) \cdot P(f,t)$.

4.1.1. Amplitude and Frequency

Changing the volume of S is trivial: given an amplification factor K, just consider the sound $(K \cdot A, F, C, W)$. Pitch-shifting is quite as much easy: given a transposition factor K, consider $(A, K \cdot F, C, W)$. The base-2 logarithm of K is the amount of pitch-shifting expressed in octaves.

4.1.2. Color and Warping

When K is the spectral response (color) of a filter, filtering (coloring) S with this filter can be done ideally like this: $S' = (A, F, K \cdot C, W)$. Warping S sounds a bit stranger, since we are not used to this parameter in music. However we are carrying out promising experiments in close collaboration with composers of electro-acoustic music. Warping is related to inharmonicity, and one can change it like this: $S' = (A, F, C, K \cdot W)$, where K is a "warping envelope". One can also perform a new kind of cross-synthesis by replacing the warping of one sound by the one of an other.

4.1.3. Time

Since all the SAS parameters are functions of time, time stretching is only a matter of scale on the time axis. For example the sound $(A(kt), F(kt), C(kt), W(kt))$ is k times shorter than S.

4.2. Combining Sounds

One of the advantage of the SAS model is its aptitude for creating hybrid sounds. One can perform many kinds of cross-syntheses only by interchanging parameters among different sounds. Figure 2 illustrates a cross-synthesis on the color parameter between S_1 and S_2. Of course one can

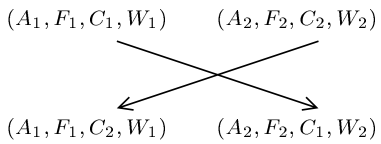

Figure 2: *Cross-synthesis by color swapping.*

also blend the parameters of different sounds. Let us realize a morphing in the SAS model. Assuming that we perceive all the parameters logarithmically, consider the blending operator:

$$\forall \alpha \in [0, 1], \quad blend(P_1, P_2, \alpha) = P_1^{(1-\alpha)} P_2^{\alpha}$$

By blending each parameters and making α vary with time from 0 to 1, one hears a morphing from the first sound to the second. But the $blend$ operator is not well-suited for voice, since it does not really care about formants. In order to perform "formant-morphing", the blending operator has to be changed for the color parameter.

4.3. Impossible Effects

Reverberation, as in additive synthesis, can not be implemented. But since SAS models (monophonic) sound sources – and not sounds reaching our ears – reverberation has no meaning in the SAS model. Echoes can not be done either, since echoing one (monophonic) sound may lead to a polyphonic sound that can not be represented as one single sound in the SAS model. Even the mixing of two sounds in the SAS model is not a sound in the model. To manipulate polyphonic sounds – that is sets of monophonic sounds – a symbolic structuration must be added on the top of SAS.

5. CONCLUSION

In this paper we have introduced the use of the Structured Additive Synthesis (SAS) model for digital audio effects. Since there is a close correspondence between the model parameters and perception, the control of the audio effects becomes simplified. Many effects thus become accessible not only to engineers, but also to musicians and composers.

While structuring the sound representation facilitates the design and control of musical sound effects, it also imposes limitations not only on the sounds that can be represented but also on the effects that can be performed on these sounds.

We are developing a sound synthesis language based on SAS, in close collaboration with composers of electroacoustic music. In order to use SAS for the whole compositional process, a hierarchical model must be designed on the top of SAS and incorporated in the language. This further structuration level makes more things possibles, like representing polyphonic sounds or performing echoing or mixing.

6. REFERENCES

[1] Myriam Desainte-Catherine and Sylvain Marchand, "Vers un modèle pour unifier musique et son dans une composition multiéchelle," in *Proceedings of the Journées d'Informatique Musicale (JIM'99)*, 1999, pp. 59–68.

[2] Myriam Desainte-Catherine and Sylvain Marchand, "Structured Additive Synthesis: Towards a Model of Sound Timbre and Electroacoustic Music Forms," in *Proceedings of the International Computer Music Conference (ICMC'99, Beijing)*, 1999.

[3] Marie-Hélène Serra, *Musical Signal Processing*, chapter Introducing the Phase Vocoder, pp. 31–90, Studies on New Music Research. Swets & Zeitlinger, Lisse, the Netherlands, 1997.

[4] Kelly Fitz and Lippold Haken, "Sinusoidal Modeling and Manipulation Using Lemur," *Computer Music Journal*, vol. 20, no. 4, pp. 44–59, 1996.

[5] Xavier Serra, *Musical Signal Processing*, chapter Musical Sound Modeling with Sinusoids plus Noise, pp. 91–122, Studies on New Music Research. Swets & Zeitlinger, Lisse, the Netherlands, 1997.

[6] Tony S. Verma and Teresa H.Y. Meng, "Time Scale Modification Using a Sines+Transients+Noise Signal Model," in *Proceedings of the Digital Audio Effects Workshop (DAFX'98, Barcelona)*, 1998, pp. 49–52.

[7] Sylvain Marchand and Robert Strandh, "InSpect and ReSpect: spectral modeling, analysis and real-time synthesis software tools for researchers and composers," in *Proceedings of the International Computer Music Conference (ICMC'99, Beijing)*, 1999.

[8] Sylvain Marchand, "Improving Spectral Analysis Precision with an Enhanced Phase Vocoder using Signal Derivatives," in *Proceedings of the Digital Audio Effects Workshop (DAFX'98, Barcelona)*, 1998, pp. 114–118.