Models of continuity between synthesis and processing for the elaboration and control of timbre structures.

J.B. Barriere, T. Potard, P.F. Balsme
IRCAM, 31 Rue St. Merri, Paris, 75004, France.

Abstract
This paper deals with the possibilities of a compositional environment that should allow a continuity between synthesis in general, and processing of instrumental sounds. The main idea of our research was to favour the elaboration and control of timbre structures through the design of a network of relations between different models of oscillations and models of resonances.

This research was conducted with the version of CHANT allowing both synthesis and processing and with the FORMES environment. We studied instruments from the different families: wind, strings, and percussion as well as voice, in order to establish a kind of dictionary of excitation and resonance terms.

This work allows one to play with static models as well as dynamic structures of transformation, working out transformations inside a single model or hybridization and transitions between different models.

Introduction
We have been studying the possibilities of a compositional environment allowing a continuity between synthesis in general, and processing of instrumental sounds. The main idea was to favour the elaboration and the central of timbre structures through the design of a network of relations between different models of oscillations and models of resonances. We wanted to emphasize the interactions between sonic materials and musical organizations by making possible a process of transformations.

This research was conducted with the FORMES language containing an implementation of the CHANT program that allows both synthesis and filtering (see figure 1). This version of CHANT is running on the PPS-100 array processor controlled by a VAX-780 computer.

The starting point of this research was a long-term project on timbre inside the Music Research Community at IRCAM, tracing its origins, like many other projects, in the developments realized by Xavier Rodet and Yves Potard with the CHANT/FORMES group since 1989 at IRCAM.

Another starting point for this research was the conceptual elaboration of "Timbre", a piece by Jean-Baptiste Barriere based on the genetic idea of unfolding materials, coming out from the same origin but evolving and being transformed in many different ways, discovering then new identities.

We were interested in developing different musical strategies around the same concept of interaction between materials and organization, two important aspects of which are timbre hybridizations and timbre interpolations (see also the parallel works of Jan Vandenheede and Jonathan Harvey, Stephen Readman and Atsuki Saitoh in these Proceedings).

Figure 1. PPS and filter synthesis on the PPS array processor.

1) Timbre and continuity between synthesis and processing
From the very beginning of computer music, composers have been interested in developing new tools for timbre composition. Most of the research done in sound synthesis is however concentrated in exploring as far as possible the advantages of one synthesis technique. Few are trying to unify these two fields by delimiting the respective domains of application of such techniques. It is also now possible to extend this field by building a bridge between synthesis and processing as soon as we are able to control them both with the same kind of rules.

The purpose is at first compositional: a composer wants to be able to think in terms of constraints and to be able to set up rules (in a very general sense, i.e. discrete points, differentiable elements of a set) for the differences parameters he decides to deal with. Related to timbre this means being able to specify different synthesis "instruments" and their behaviour in a variety of contexts. To achieve this, we need to understand something about the functioning of each instrument in an intuitive a way as possible. One point of view is to privilege sound production as it can be inferred from perception. This gives direct care for the composer's imagination.

ICMC '85 Proceedings
This is why we decided to privileges in our research a kind of continuing hypothesis. The imagination needs to proceed to relations between objects, extrapolating the unknown from the known, offer by a very similar process to those of improvisation or extrapolation.

We therefore chose to work on timbre starting with the most general reference to the composer: the instruments of the orchestra. We considered them as points in a plan, that can be generated by a number of parameters, in different modes. The transfer from one feature of an instrument to another instrument is going further than the idea of setting up a collection of "unbridled objects". It allows the setting up of a fact inside which the composer can more after having designed the proper timbre of the instruments, working with in his composition. These movements become parameterizable. The idea of translation becomes interesting not only because it involves a directability (you are going from one point to another), but because it makes possible the discovery of intermediate points (you create by yourself points on the path), and you may as well decide to use these points as entities by themselves, like some other instruments. These hybrids, incidentally, may be thinkable in the imagination of the musician when proceeding with the reference to the production-experience axis as discussed before, but unavoidable in the material world. They will nevertheless be most of the time understandable as such by the listener.

In this process the fact that we always keep a balance between intuition and calculation provides a guarantee on the result, for the listener, which is then the source for the composition itself. In that very sense the sonic material can support the organization. In a similar way ideas about organization can suggest the elaboration of a sonic material. Starting from the voice, one can imagine superimposing the instruments hypothesis, in which an alternating process and consonant on a sound source by a complex filtering process. This will produce a voiced sound. In that example we may have a very complex organization process to an object that has a very complex material, but a very simple organization. The result is a complex sound that has a very simple structure. Typically this idea can be transposed to most resonant instruments especially to percussive instruments, such as for instance in a different flame to played strings.

Let us clearly the filtering method we used here. This is very different from applying only the dynamic profile of an instrument to another instrument. We model the temporal behavior of the instruments by controlling mainly its temporal spectral evolution, and then apply it to control the order filters in terms of frequency, amplitude, and bandwidth. In the case of the voiced sound, this is like replacing the vocal chords and the noise source in the vocal spectrum by another source of excitation, a sound, placed in the vocal tract.

The same examples can be projected, of course, to many different excitations and many different resonant poles, provided we can model them in a proper way. This depends on the complexity of the physical production of the sounds and on our ability to model them.

We started to systematize that idea over a large instrumentarium, trying to cover the largest possible number of families, and it was always possible with very good results.

At first, we studied instruments belonging to different families (e.g., strings and percussion as well as voices). In order to establish a kind of dictionary of forms of excitations and resonances in the basis of our work about hybrids and transitions (see figure 2).

2) The modeling process: some basic models

2.1) Methods for modeling

Before describing the elaboration of the models, let us throw light on the distinction we made between excited and resonant. This is a pragmatic and operational distinction, based on perceptual criteria, and well suited for the TDF synthesis technique in which excitation and resonance are directly linked although separably controllable.

An excitation can mean a recorded sound or the simulation of a physical excitation or a noise source. A resonance is a spectral envelope defined in terms of formants or filters, depending on the nature of the excitation: formants when doing synthesis, filters when doing processing. It is possible to combine different excitation models. And it is also possible to chain together different resonance models. For example, synthesizers that simulate an abstract control of those elements. One can play with static models as well as with dynamic structures of transformation, working out transformations within a single model or transitions between different models.

When elaborating a double-base model, for instance, we considered it a classical way that the excitation is the player's finger or the blow, whereas the resonator is the set composed of the string, the bridge and the body of the instrument. A distinction between the sets (finger or blow setting) and (instrument/body) could have been chosen, but it appeared to us as a less pertinent choice. In a physical object, the most salient resonances are the longest ones — those of the string in the double-base case — and they should be in the most relevant for the elaboration of a resonance model. Ideally, the model to be built should reproduce the pure transfer function of the resonator. Practically, we deliberately chose to derive from classical methods of instrumental acoustics, and to start the analysis with "musical" sounds. The latter inevitably bear the stamp of a
given mode of excitation, that we expected to eventually fall in with in the model. The chosen sounds were produced by a simple impulse: a slap on the mouthpiece for the tube, a hammer for the double-bass for instance. The following analysis methods were used:

* Long-term FFT: typically, we chose windows 5 to 10 second large: we obtained the center frequencies of the resonances, their mean amplitude, and the bandwidths when the resonances were not too sharp, as it was the case for the tube.
* Phase-Vocoder: the channels for the analysis were roughly centered on the harmonic partials, and did not overlap with one another.
* The temporal evolution of the amplitude in a given channel yields:
  - the amplitude of the resonance at the attack
  - the bandwidths of the sharp resonances.
* In addition, the periodic amplitude modulations (beating) lend themselves to subtle interpretations. The frequency evolution allows:
  - a confrontation with the FFT results
  - an integration of temporal rules resulting in dynamic models
  - an eventual distinction between close partials.

2.2) Three representative models

We worked to obtain several basic models, from which we shall describe more representative ones: the tubular bell's, the tube, and the double bass.

2.2.1) Tubular bell

The FFT analysis of a knocked bell tone was used to localize the partials in the frequency domain, with the phase-vocoder analysis, we determined the partial bandwidth, their initial amplitudes (which in fact depend on the mode of excitation), the duration of their attack (which is controlled in FOF synthesis by the parameter 'ten'). The analysis of the beating gave the amplitude and the frequency of some additional partials.

2.2.2) Tube

This instrument is obviously more complex; it has a variable tube length, the exciting cause can be a slap, a buzz, a knock, a breath... The resonator class is a compounded one: the mouthpiece and the body of the instrument are in series and have different transfer functions, which, when multiplied, result in the global transfer function. The contribution of the mouthpiece is taken into account in the model, but we supposed that it could be reduced to a global spectral envelope with relatively large bandwidth (and thus a less important perceptual salience), superimposed on the transfer function of the body of the instrument (see figure 3).

Through an FFT analysis of a slapped tone, we obtained the frequencies of the first fifteen partials, their bandwidths and their amplitudes (see figure 4). The latter integrate the on-top of the mouthpiece, which has been determined in another way.

The parameters for higher partials are set through a simple rule, in the FOFXEX program controlling the synthesis. Another rule changes the parameters of the resonances according to a chosen length of the tube. At least, FFT analysis of various parts of a heavy tone yielded an additional rule in the FOF model, which affects the spectral slope, and which could typically be useful to different contexts in this work (see figure 5). This can be considered as a modeling of the excitation produced by the player's lips, that could also be transposed in the filter model.

Figure 3. Spectrum of buzz in a mouthpiece separated from the body of the instrument.

Figure 4. Spectrum of a slap on the tube mouthpiece.

ICMC '85 Proceedings
8.2.3) Double-bass. We used a phase-locked analysis of a piano tone produced on the first (open) string. We directly modeled the transfer function of the set [string/body]. Let us describe the main features that the analysis revealed, with the amplitude and frequency evolution of the channel corresponding to the third harmonic of the fundamental tone (c3). In figure 9, one can observe that the bandwidth, which is linearly related to the inverse of the amplitude "fatality" decreases with time. One can also clearly distinguish two amplitude modulations. From the slow one, we can deduce a simple modeling of the bandwidth decreasing, when considering that the modulation is produced by two close resonances with different bandwidths and amplitudes. The period of the modulation yields the distance in time of the two resonances. The more damped resonance is the louder one at the attack; after the biggest modulation—which corresponds to equal amplitudes for the two resonances—the one occurred, the loudest resonance dominates; a slight shift of frequency can then be observed. This calls to mind G. Wrench's model of a piano string, where two resonators corresponded to two cortical and transversal modes of vibration. The fast and deep modulation, which is easily noticeable at the attack for several partials, can be modeled by adding a third resonance with fast damping and located about (0 Hz) below the main partial. We have not been able to now to give a satisfactory physical interpretation of this phenomenon.

![Figure 8: Amplitude and frequency evolution in the channel corresponding to the third harmonic of an all double-bass tone (pizzicato).](image)

ICMC '85 Proceedings
Towards hybrid models

For musical reasons, when starting the specific work with the tube, we came to be interested in the instrumental situation where the player is both playing the tube and singing to it. This is already an hybrid, a natural one. Musicians and acousticians know well that singing and playing simultaneously on a brass instrument causes the advent of an additional set of partials which do not belong to one or the other source, but which is the result of a kind of ring modulation effect between the 2 sources, and which is the sum and difference of both source partials.

We were seduced by this phenomenon because it gave an interesting role to work with inharmonic sounds starting from periodic or quasi-periodic sources. Again this provided us a proper path in our desire for continuity.

Once the tube had been properly modeled we started hybridized experiments. The first one we did was to take the sound of a tube player (Gerard Touriet) and help a lot in the understanding of the instrument) breathing heavily (in a similar way as when playing) to the mouthpieces separated from the body of the instrument. We then put this recorded sound after digitization as an external source through the tube model controlling the fingers. The result was immediately a very realistic tube sound. The next step was to replace this external source with the tone source. The result was of course less essay but anyway still convincing. We then processed through the tone model a singing voice, synthetic last time, after having modeled properly, within the "instrument" itself, the ring modulation with an additive synthesis model realized in PDP. The inputs to this real modulation process were simply the fundamental frequency values of both pseudo-sources. Again the workfell as expected.

Figure 9. Synthetic sung vowel (a). Tube low and changing tone (b). Sung vowel filtered with the tube (c). Tube filtered by the sung vowel (d).

ICMC '85 Proceedings
The first line at this point was to try out the reverse process. We originally proposed this technique to improve the accuracy of a particular model in a specific domain. However, during the early stages of development, we faced several challenges. This led us to reevaluate our approach and consider alternative strategies. Ultimately, we decided to focus on refining our existing methods rather than adopting a radical new approach. Moving forward, we plan to continue experimenting with various techniques to optimize performance.

As we refine our methods, we are also paying close attention to potential ethical implications. It is crucial that our work does not undermine the trust of our users. We are committed to conducting thorough research to ensure that our models are both effective and responsible. We will continue to collaborate with experts from diverse fields to gain a broader perspective on the impact of our work.

In conclusion, while we have encountered some setbacks, we remain optimistic about the potential of our approach. We are confident that by persistently pushing the boundaries of what is possible, we will be able to make significant contributions to the field. We encourage others to join us in this pursuit and look forward to sharing our findings with the wider community.