An Abstract Approach to Music[1]

Hans G. Kaper

Mathematics and Computer Science Division, Argonne National Laboratory

Sever Tipei

School of Music, University of Illinois

Abstract

The notion of formalized music implies that a musical composition can be described in mathematical terms. In this article we explore some formal aspects of music and propose a framework for an abstract approach.

In his fundamental treatise on *Formalized Music* [1], Iannis Xenakis shows how a musical composition can be structured on the basis of mathematical ideas. His sieves are essentially implementations of set- and group-theoretical concepts, which can give structural coherence to a musical composition; his stochastic music is based entirely on notions of probability theory; and the idea of a composition as a book of screens points to an interesting connection with the branch of mathematics known as analysis. Many compositions based on these ideas, by Xenakis as well as others, testify to the validity of a formal approach to music.

The notion of formalized music implies that a musical composition can be described in mathematical terms. In its most extreme form, formalized music is an algorithm—a description of a musical composition as a sequence of steps in a computational procedure. The composer creates the algorithm, and the performer renders the algorithm into aural events. Since computers understand only formal languages, computer music is necessarily formalized music.

Our purpose in this article is to create an abstract framework that enables an algorithmic approach to music. We do not give an algorithmic approach *per se*, and many details have yet to be filled in, but we have at least the beginning of an abstract framework. The article consists of three parts: Section 1 presents basic concepts and terms, Section 2 the formal framework, and Section 3 a summary and conclusions.

[1]This work was supported by the Mathematical, Information, and Computational Sciences Division subprogram of the Office of Computational and Technology Research, U.S. Department of Energy, under Contract W-31-109-Eng-38.

1 Basic Concepts

The compositional process is based on the assumption that aural events can be ordered in time. While we think of a musical composition as an abstract object, its realization in an actual reading is a time-ordered event. Without time, there is no music. This observation, although obvious, is basic to an algorithmic approach to music and music composition as proposed in this article.

1.1 Trajectories in Sound Space

In abstracto, a musical composition stands for a trajectory (or a set of trajectories, in the case of aleatory music) in a space of aural events. A trajectory is simply an ordered set of points. When the composition is realized, the trajectory is traversed in the course of time, starting from a given position until the end of the piece is reached. This image suggests that a musical performance is a *dynamic event*, which evolves in the space of aural events. (We sometimes use the shorter term "sound space" to denote the space of aural events.)

The compositional process presents the composer with the opportunity to control the structure, if not the details, of the trajectory and thereby the nature of the composition. The control takes the form of an algorithm—a set of rules governing the evolution of the objects that, together, constitute the musical composition. Formalizing music thus means defining the objects and seeking those attributes of the objects that define the trajectory in sound space. In this article, we present a tentative model for this process.

Remark. Evolution is not the same as change. Just as a mathematical function can have a constant value on its domain, the aural "value" (perception) of a composition can be constant over time. A static drone is still a dynamic event, although there is no change in the aural perception over time. From this perspective, the phrase "The evolution is nil" [1, Chapter II, p. 50] does not make sense.

1.2 Partials, Sounds, and Compositions

The first concern in a formalization of music must be the objects that constitute a musical composition.

As Western musicians, we are trained to think of music in terms of the notated score. Consequently, we conceive of a musical composition as a collection of sounds, each sound being identified with a dot or oval in the score. We consider sounds (including rests) the essence of music.

One can argue whether the concept of a sound offers the best starting point for an algorithmic approach to music composition. After all, one can analyze a sound in terms of its partials, so maybe a partial is a more fundamental object. Of course, a partial is only a construct—we don't perceive the individual partials in a sound. But we do notice a change in the sound when one of its partials is modified. Furthermore, a partial is the basic unit in additive sound synthesis.

On the other hand, one could take groups of sounds, such as fragments and cells, as the objects of choice and try to build a formal structure of music out of these elementary building blocks. In fact, one could push this idea even further and declare that the entire composition is the truly fundamental unit of music.

One faces a similar issue in physics when one tries to describe matter in terms of elementary objects. Are atoms more basic than elementary particles (which one may not even be able to observe directly) or than macroscopic objects? The analogy is, in fact, not too far fetched and points to a possible strategy to resolve the issue.

Physics is the science that looks for unifying principles in nature—principles that are sufficiently abstract that the resulting formal description embraces natural phenomena on very different scales. Quantum mechanics is more abstract than classical mechanics, and each describes matter in a well-defined regime. The descriptions use entirely different objects (waves, particles) but become indistinguishable on the macroscopic scale.

If a musical composition represents a trajectory in a space of aural events, then the same is true for the sounds that make up the musical composition and the partials that make up the sounds. But partial, sound, and composition evolve on very different time scales. The characteristic time for a partial is determined by the frequency of the pure tone, typically on the order of 10^{-3} seconds for a tone in the audible range of the spectrum. The characteristic time for a sound, on the other hand, is determined by the frequency of a modulating wave and is therefore on the order of 10^{-1} seconds. If we consider an entire musical composition, we measure time in minutes and seconds, so the characteristic time is on the order of 10 seconds. Roughly speaking, the time scale increases approximately by two orders of magnitude at every step.

This observation suggests a hierarchical ordering of partials, sounds, and compositions as sketched in Figure 1. Is it possible to develop a formal structure of music that covers

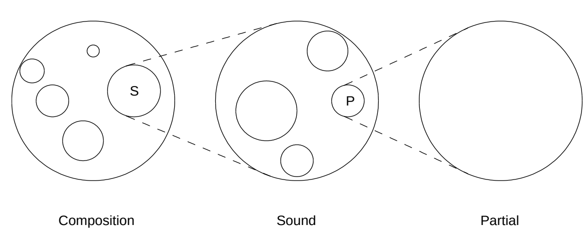

Figure 1: Hierarchical model of partials, sounds, and compositions.

the structure of the hierarchy as well as that of the individual elements in the hierarchy? By focusing on this question, music theory becomes the science that looks for unifying principles in the space of aural events.

In principle, we could include collections of sounds such as fragments and cells as additional objects in the hierarchy of Figure 1. These objects would fall somewhere between sounds and compositions. However, we prefer not to do so, for various reasons. First, we wish to keep the framework simple. But, more important, the time scale for sounds, fragments, and cells are comparable. This does not necessarily mean that they are the same, but they are sufficiently close that we can treat sounds, fragments, and cells as similar objects. Also, we think of an object as a representative of a class of similar objects. The objects in the class have the same attributes, they obey certain relationships, and they can be subjected to the same operations. For example, a sound has a certain loudness, we can order sounds according to their (perceived) loudness, and we can change the dynamics of a sound. In this sense, fragments and cells are similar to sounds, but fragments, cells, and sounds are distinct from an entire piece. Finally, there are cultural reasons to exclude fragments and cells from the hierarchy, as we explain in the next section.

The hierarchy of Figure 1 suggests a natural inheritance scheme of attributes and operations. Thus, we have the beginning of a model for an algorithmic approach to music.

Remark. The image of classes of objects within classes of more general objects extends in an interesting way. A musical composition represents, as we have seen, a trajectory in a space of aural events. This space is full of aural events; some correspond to musical compositions, but most do not. We can turn the space of aural events into an object by defining relations among the objects and ways to change their attributes. But this object is just one instantiation of a class of much more general objects, namely, the class of all events. There are other instantiations of this class, for example, the class of visual events, or the class of all two-person games. Each instantiation inherits the attributes of the larger class of more general events. Thus, a trajectory in the space of aural events can be related to an instantiation of a two-person game as realized in a duel. Remarkably, this idea was anticipated by Xenakis and implemented in his composition *Duel* [1, Chapter IV, p. 113].

1.3 Cultural Influences

The definition of an object such as a composition or a sound needs to be sufficiently abstract and, we claim, free of stylistic constraints. Can we infer its attributes from a score?

Consider the case of a sound. The notes in a score characterize a sound by its primary attributes: starting time, duration, and pitch. Other features in a score that relate to the attributes of a sound can be inferred from the markings for dynamics (loudness), instrumentation (timbre), and articulation (shape), maybe even distance cues and instructions for movement in space.

Is it possible to use these attributes in a formal approach to the concept of sound? We don't think so. The many attempts to define a sound's timbre bear witness to the inherent pitfalls of such an approach. The problem is that the attributes are defined at a relatively low level of abstraction—for example, with reference to the instrument in the case of timbre. A low level of abtraction leads to a plethora of definitions, with the risk of inherent contradictions and redundancy resulting in nonsense.

Also, a formal approach to music must be free of cultural bias. The traditional terminology and concepts defining a sound by its primary attributes starting time, duration, and pitch are tailored to Western music. Other musical cultures consider duration and pitch important qualities too, but the primacy of the latter is sometimes abrogated. In African drumming, for example, rhythm and timbre alone are the ingredients of music, and in some traditional music of the Far East, timbre takes precedence over pitch. It is fair to say that, even in some recent Western music, the traditional way of thinking about music and sounds has lost much of its relevance.

Cultural bias is one of the reasons that we included only partials, sounds, and entire compositions in the hierarchy of Figure 1. Concepts such as fragments and cells can, in principle, be viewed as objects, but they are culturally determined and reflect traditional Western, rather than universal, notions of music.

1.4 Toward a Definition of Musical Objects

We claim that a high level of abstraction is fundamental for any formal approach to music. Moreover, the abstraction must be free of stylistic constraints. What then *is* a sound, or a composition?

We claim that, at their most elementary level, sounds and compositions are the manifestations of complex audio waves. Hence, we take the *audio wave* as the defining characteristic of a musical object—be it partial, sound, or composition. The audio wave has two aspects, one physical (the variation of the ambient air pressure), the other psychophysical (the process that translates these vibrations into a perception of the sound). The algorithmic definition of the object must accommodate both aspects.

As we have seen, time is the truly independent variable. It governs any aural experience, and we cannot control it. Accordingly, we consider the complex audio wave as a function of time with values in sound space. This may seem like a verbal trick, but it is a significant step in the process of abstraction. The description of sound space has now become an integral part of the definition of a musical object.

Sound space is a multidimensional vector space. Its elements are functions of many variables (degrees of freedom). Once values are assigned to the variables, the position of an object in sound space and, hence, its perceived qualities are completely determined.

Remark. The idea of a musical composition as a mapping from the time domain into sound space is closely related to the concept of a musical composition as a *book of screens*, an idea first proposed by Xenakis [1, Chapter II, p. 50]. Unlike Xenakis, however, we view time as an independent variable, not as another degree of freedom in sound space. In Xenakis's terminology [1, Chapter VI, p. 159], *temporal structure* is independent of *structure outside-time* and *structure in-time*. We return to this point in Section 2.2.

2 Formal Structure

We now formalize some of the concepts discussed in the preceding section. The formal structure is based on the hierarchy of partials, sounds, and compositions sketched in Figure 1. Partials, sounds, and compositions are defined as *objects*, which have attributes, obey relationships, and are subject to certain well-defined operations.

2.1 Compositions, Sounds, and Partials as Objects

The universal object in the space of aural events is the *audio wave*. Special cases are partial waves corresponding to pure tones, sound waves corresponding to sounds, and complex audio waves corresponding to entire musical compositions. Partial and sound waves are like threads floating in the space of aural events, which are woven into the trajectory of a musical piece by the composer. This image suggests how to formalize the corresponding objects.

The object of a musical composition is a *complex audio wave*. We denote it generically by the symbol W. Two of its attributes are its starting time ($T_{w,0}$) and its duration (T_w). (The subscript w stands for "wave.") Thus, a musical composition (or its representation, the complex audio wave) is described by the set of all values $W(t)$ on an interval of length T_w beginning at $T_{w,0}$,

$$W = \{W(t) : t \in [T_{w,0}, T_{w,1}]\}, \quad \text{where } T_{w,1} = T_{w,0} + T_w. \tag{1}$$

Notice the difference between W and $W(t)$: W is a trajectory (a set of points) in sound space, whereas $W(t)$ is a single point in sound space, namely, the point on W associated with a particular value of time, t.

This description of a composition as a complex audio wave is independent of the time the piece actually starts or ends: both $T_{w,0}$ and T_w are attributes (degrees of freedom), to which we assign a value when we realize the piece. Since both are independent of time, they are *static* attributes.

The complex audio wave itself is the superposition of its constituent sounds. Hence, its value at any moment t in the interval $[T_{w,0}, T_{w,1}]$ is given by an expression of the form

$$W(t) = \sum_{i \in I_w(t)} S_i(t), \quad t \in [T_{w,0}, T_{w,1}]. \tag{2}$$

Here we encounter another attribute of the object W: I_w, the set of indices of all sounds in the audio wave; $I_w(t)$ is its value at time t, and the sum extends over all sounds that are "active" at time t. The ith sound contributes a value $S_i(t)$ to $W(t)$. The sound S_i may be a single partial or, more generally, a superposition of partials. Note that I_w is a *dynamic* attribute of the wave; its value may vary with time. In general, this variation occurs on a time scale that is characteristic for the composition.

We realize the composition by assigning values to its attributes. The values are real numbers in the case of static attributes and functions in the case of dynamic attributes. In the latter case, we specify the attribute's shape (envelope function) and size (maximum value).

The ith sound S_i in Equation (2) is an instantiation of the class of sounds. The definition of a sound is analogous to that of a composition. A sound S of duration T_s is the set of all its values $S(t)$ on an interval of length T_s beginning at $T_{s,0}$,

$$S = \{S(t) : t \in [T_{s,0}, T_{s,1}]\}, \quad \text{where } T_{s,1} = T_{s,0} + T_s. \tag{3}$$

Here, $T_{s,0}$ and T_s are (static) attributes of the sound object, to which values are assigned when the piece is realized.

A sound is the superposition of its constituent partials, just like a composition is the superposition of its constituent sounds. Hence, the value of a sound S at any moment t in the interval $[T_{s,0}, T_{s,1}]$ is given by an expression of the form

$$S(t) = \sum_{j \in I_s(t)} P_j(t), \quad t \in [T_{s,0}, T_{s,1}]. \tag{4}$$

The symbol I_s denotes the set of indices of all partials in the sound S; $I_s(t)$ is its value at time t, and the sum extends over all partials that "actively" contribute to the sound. The jth partial contributes a value $P_j(t)$ to $S(t)$. The index set I_s is a dynamic attribute of S; it varies in time, but the variation occurs generally on a time scale that is characteristic for the sound.

Finally, the jth partial P_j in Equation (3) is an instantiation of the class of partials. A partial P of duration T_p is again the set of all its values $P(t)$ on an interval of length T_p beginning at $T_{p,0}$,

$$P = \{P(t) : t \in [T_{p,0}, T_{p,1}]\}, \quad \text{where } T_{p,1} = T_{p,0} + T_p. \tag{5}$$

A partial being the elementary object from which the other objects (sound waves, complex audio waves) are built up, we identify it with a sinusoidal wave with amplitude a, frequency f, and phase ϕ,

$$P(t) = a(t) \sin(2\pi f(t)t + \phi(t)), \quad t \in [T_{p,0}, T_{p,1}]. \tag{6}$$

When the amplitude, frequency, and phase are constant in time, Equation (6) represents a segment of a pure tone. In practice, at least the amplitude will vary with time, because the suppport of the partial (that is, the closure of the set of t for which $P(t) \neq 0$) must fit in the interval $[T_{p.0}, T_{p,1}]$. But in principle, all three variables (amplitude, frequency, and phase) represent dynamic attributes of a partial, which may vary on a time scale that is characteristic for a sound; $a(t)$, $f(t)$ and $\phi(t)$ are the values of a, f, and ϕ at time t.

The choice of a sinusoidal wave as the fundamental wave type is convenient but not necessary; other wave types, such as splines and wavelets, do just as well. The main criterion is that the partials form a complete set of basis functions in sound space.

In some instances, it may be more convenient to think of a in the definition of a partial as a relative amplitude measured, for example, with respect to the amplitude of the fundamental in a sound. One can then incorporate a dynamic scaling factor in the definition of a sound, or even in the definition of a composition, and deal more easily with issues of (perceived) loudness.

2.2 Attributes

The abstract framework set up in the preceding section enables us to give a more precise meaning to most, if not all, concepts commonly encountered in music.

First, a few remarks about time. Xenakis observes that musical analysis and construction may be based on three concepts: *structure outside-time*, *temporal structure*, and *structure in-time*; see [1, Chapter VI, p. 159]. We claim that these concepts can be given a more precise meaning in the present context.

Recall that we have identified a musical composition with a complex audio wave. The wave is represented by the symbol W and defined in Equation (1). The variable t inside the braces is a parameter, which has no particular interpretation yet. The structure of W is therefore a "structure outside-time." The wave W has two static attributes: $T_{w,0}$, its beginning time, and T_w, its duration. When we realize the composition, we assign values to these attributes. Such an assignment assumes the existence of a physical time line, complete with a point of reference and a given unit of time. Xenakis's "temporal structure" is the structure of this time line, which is that of a partially ordered set. When we make the assignment, we map t onto physical time, and the structure of W becomes a "structure in-time." Thus, the formalism in our framework does not change; only the interpretation of the symbols changes.

The formal framework of the preceding section was inspired by the observation that, in a musical composition, different objects (a composition, a sound, a pure tone) evolve on significantly different time scales. The observation led us to introduce the hierarchical structure sketched in Figure 1. Yet, the different scalings are not evident in the expressions given in the preceding section; they all involve the variable t, and the dynamic attributes of the composition, sound, and partial objects all seem to evolve on the same time scale. But this is only seemingly the case; it is easily fixed if we introduce multiple time scales.

First, introduce the parameters $\varepsilon_{s,p}$, the ratio of the characteristic time for S to the characteristic time for P, and $\varepsilon_{w,s}$, the ratio of the characteristic time for W to the characteristic time for S. Measuring t on the time scale of a partial, we define the dynamic attributes of a sound as functions of a slow time $t_s = \varepsilon_{s,p}t$ and the dynamic attributes of a composition as functions of the even slower time $t_w = \varepsilon_{w,s}\varepsilon_{s,p}t$. As $\varepsilon_{s,p}$ and $\varepsilon_{w,s}$ are both small, typically of the order of 10^{-2}, t must change by two orders of magnitude to compensate for the smallness of $\varepsilon_{s,p}$ before a measurable change of t_s is obtained, and by another two orders of magnitude to compensate further for the smallness of $\varepsilon_{w,s}$ before a measurable change of t_w is obtained. The parameters $\varepsilon_{s,p}$ and $\varepsilon_{w,s}$ enable us to establish a structure among the dynamic attributes of a composition, its sounds, and its partials.

As a further step in an abstract approach to music, we would encapsulate the objects and their attributes together with the functions to manipulate the objects in a *class*. A class is a highly abstract concept, most conveniently thought of as a blueprint. Out of a blueprint, a builder can build a house. Out of a class, a composer can create an object. One class can be used to make many objects of the same class.

By way of example, we turn a sound object into a sound class by introducing the sound's *perceived loudness* as one of its attributes and augmenting the definition of the sound object with an algorithm to compute its perceived loudness. The algorithm must embody all the steps necessary to compute the loudness, irrespective of the composition of the sound and irrespective of the distribution of the partials in the sound. Part of such an algorithm will be based on the results of psychoacoustic research.

If a composition is based on stochastic elements, the composition class will contain a random number generator. The sound class inherits the attributes of the composition class, so all the partials in a sound will have access the same random process—a desirable feature if the sound is to be perceived as a sound, rather than as a collection of partials.

A class incorporates more than just the objects. It places the objects in an appropriate context, defines relationships, and forms a structure, which can unify a musical composition.

3 Summary and Conclusions

In this article we have outlined a formal framework for an abstract approach to music and music composition. The model is formulated in terms of objects that have attributes, obey relationships, and are subject to certain well-defined operations. The motivation for this approach uses traditional terms and concepts of music theory, but the approach itself is formal and uses the language of mathematics.

The universal object is an audio wave; partials, sounds, and compositions are special objects, which are placed in a hierarchical order based on time scales. The objects have both static and dynamic attributes. When we realize a composition, we assign values to each of its attributes: a (scalar) value to a static attribute, an envelope and a size to a dynamic attribute.

A composition is then a trajectory in the space of aural events, and the complex audio wave is its formal representation. Sounds are fibers in the space of aural events, from which the composer weaves the trajectory of a composition. Each sound object in turn is made up of partials, which are the elementary building blocks of any music composition. The partials evolve on the fastest time scale in the hierarchy of partials, sounds, and compositions.

The ideas outlined in this article are being implemented in a digital instrument for additive sound synthesis and in software for music composition. A demonstration of some preliminary results has been submitted by the authors for presentation at the conference.

References

[1] Xenakis, I., *Formalized Music, Thought and Mathematics in Music*, revised edition, Pendragon Press (1992)