An Abstract Approach to Music[1]

Hans G. Kaper
Mathematics and Computer Science Division, Argonne National Laboratory
Sever Tipei
Computer Music Project, 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.

An instrumental or vocal ensemble generates a complex audio wave - variations of the air pressure which make the eardrum vibrate. At a performance, the output of instruments, voices, and possibly loudspeakers, is mixed with noises from the audience and augmented with the reflections from the surrounding surfaces.It is truly remarkable that, despite all these extraneous additions, the ear and the brain are able to discriminate and identify the musical output with precision and apparent ease.

1 Levels of abstraction

Complex audio waves are traditionally defined at a relatively low level of abstraction. The terminology and basic concepts underlying the notion of a sound are tailored to Western music of the past three centuries. They are not very useful for a significant body of recent works and for some non-Western music. Our system of score notation reflects the same bias and, in turn, influences the way we compose.

1.1 Sound parameters

Western musicians are trained to think in terms of the notated score. Accordingly, they view a piece of music as a collection of individual sounds represented by dots and ovals and characterized mainly by start time, duration, and pitch.

The music notation system we use today is precise in defining time and pitch. It also allows for the relatively accurate transcription of dynamics and articulations. Timbre, on the other hand, is defined only with reference to the means of production: a particular voice or instrument. In the case of extended techniques, clever symbols or explanations convey a reasonable amount of information, but they seldom provide an exhaustive description. Moreover, when the waveform itself is composed from scratch, as it is in software synthesis, both the complexity of what we call "timbre" and the failure of the terminology become apparent.

We propose a formal definition of a sound that is based on the observation that a sound is the manifestation of a complex audio wave. The audio wave has two aspects, one physical

---

[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.

(the variation of the ambient air pressure, which makes the eardrums vibrate), the other psychophysical (the process that translates these vibrations into a perception of the sound). The definition we propose is sufficiently abstract that it accomodates both aspects.

Following Xenakis, we view a complex audio wave as a dynamic event that evolves in a multidimensional vector space ("sound space"). Unlike Xenakis, however, we view time as an independent variable, not as a degree of freedom in sound space. Sound is thus a vector-valued map from the time domain to sound space. The description of sound space is an integral part of the definition of a sound.

A sound space is spanned by a complete set of independent vectors, each of which is associated with a degree of freedom of a sound. These vectors constitute a basis in the sound space. A basis is not unique, but a representation of a sound with respect to a given basis is.

Deciding which are the basis vectors of the vector space or, in other words, which aspects of the sound are going to be taken into account is of consequence. The old terminology reflects traditional choices and is incomplete and often misleading, especially when it comes to timbre – itself a complex conglomerate of several degrees of freedom. At the same time, computer sound synthesis techniques allow a precise and fine manipulation of each of these components.

We claim that only choices consistent with a mathematical description of a complex wave result in a workable definition of a sound. The mathematical model we propose is based on the identification of an appropriate basis of the vector space in which we represent sounds ("sound space").

1.2 Basic Formalism

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 $T_{w,0}$,

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

Note the difference between W and $W(t)$: W is a trajectory (a set of points) in sound space, whereas

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.

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 (**??**) 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

$I_s(t)$ 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 (**??**) 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 (**??**) represents a segment of a pure tone. In practice, at least the amplitude will vary with time, because the

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.

1.3 Modifiers

Most musical and environmental sounds are actually quite complicated. They have a beginning (attack) and and end (decay), hence a finite duration; their frequency seldom remains constant, and their amplitude varies even more. Such time variations may include tiny and slow oscillations around an average frequency, arbitrary "sound bends," glissandi, and vibrato or frequency modulation (FM). With respect to amplitude variations, one could distinguish envelopes covering the life-span of partials, slower increases and decreases, sometimes over many sounds (crescendo/descrescendo), and tremolo or amplitude modulation (AM).

We can extend this concept by including some of the following:

Transients, or sudden "spikes," brief disturbances in either frequency or amplitude characterized by magnitude and frequency of occurrence. There is no reason why they should be restricted only to the attack portion of a sound as in the case of acoustic instruments.

Phase modulation is barely distinguishable from frequency modulation, and the conventional wisdom among musicians is that it does not have a significant effect on the sound. However, phase is essential in determining the location in space of a sound source and by controlling the phase one can stimulate the illusion of a spatially localized source.

Reverberation can be included in the mathematical model by adding sound reflected from surrounding surfaces to the original signal. The reflected sound can incorporate the effects of hall size and wall coverings by appropriately chosen delays and attenuation factors. The superposition of direct and reflected waves is an additive process similar to the superposition of sounds in a chord. The same effect also suggests fundamental composition procedures, such as imitation and heterophony.

The distribution of sound sources in space, the domain of what more recently has been called "acousmatic" music, is not yet included in our model, primarily because of the lack of a good mathematical model. A true spatial reproduction of sound will have to be three-dimensional and will probably have to include phase cancellations, HRTFs, or both.

In the case of traditional instruments, such "modifiers" act globally, affecting all the partials of a sound in more or less the same way. At a more abstract level, however, each partial can be modified individually. The spectrum of partials in a sound can vary dynamically, some envelopes may return to zero while others are still active, and a new partial may bloom long after the beginning of the sound. A wave's mathematical description allows for the total tion allows for the total independence of all its partials under all operations.

1.4 Lively randomness

Electronic sounds, especially those created in the early days, can be easily recognized as such because, unlike those produced by acoustic instruments, they lack a host of small fluctuations in the details of their makeup. Perfectly steady frequencies, exact durations, a vibrato made out of periodic oscillations, etc. give away their origin. One way of producing livelier, less artificial sounds is to introduce small deviations from the intended values and allow these deviations to change in time. By increasing the complexity of electro-acoustic sounds and the amount of information they deliver one can bring them closer to their "natural" counterparts. In our model, we introduce random functions to control such fluctuations and avoid the occurrence of recognizable patterns.

2 Time scales

The "modifiers" listed in Section 2.2 are easily recognized and each could be regarded as a separate quality of sound; often, they are considered as such for compositional purposes. However, formula (1) views sound as a dynamic process in a three-dimensional space. The degrees of freedom are amplitude, frequency, and phase. A closer inspection reveals that the dynamics evolve on three different time scales: one determined by the audio range, and two others which are determined by the "modifiers." This is shown schematically in Table 1.

Identifier Scale (in seconds) ————————————————————————- Wave/partial frequency ———————————————————————— Stable spectrum Amp. envelope duration of sound AM, FM, Freq. fluctuations Transients

AM, FM, Freq. fluctuations Transients ————————————————————————- Sound bend/gliss Reverberation duration of groups Spatial movement of sounds Cres/Descresc —— ——————————————————————-

2.1 Two observations

The definition of time regions within which certain sound qualities are active expands on an idea first proposed by Karlheinz Stockhausen in his article "...how time passes..." over four decades ago. Stockhausen noticed that duration and pitch belong to the same continuum: they both are "alterations in an acoustic field (...) time-intervals of varying magnitude" [Stockhausen, 1957].

Composition and sound synthesis overlap and are intertwined in an organic way. The elements of third time region could belong to either area depending on the granularity adopted in order to look at a unique, uniform, and continuous process. When writing for acoustic instruments, the task of composing "timbre" is confined to the choice of the sound source, over which comparatively little control can be exercised. But electro-acoustic music technology allows a much deeper and finer intervention in the details of a musical piece.

3 Defining sound

The expression (1) describes a wave which can extend arbitrarily in time. Partials may be active or not at various times t, and some of them might share certain features: start time, amount of AM, FM, or reverberation, while their frequencies may be in a harmonic relationship.

Traditional music regards groupings of partials with such common features as stable associations. They are identified as "sounds." The mathematical expression allows for them, but does not require them. They simply are a convenient way to show that certain global operations

The mathematical expression allows for them, but does not require them. They simply are a convenient way to show that certain global operations apply to a collection of partials at the same time.

A system of notation describing the behavior of each partial will be incredibly cumbersome, and also useless in the case of orchestral music, where partials share qualities and "modifiers" by default. On the other hand, controlling the music at this level of granularity and being specific about it is necessary when the composition process extends to finer details.

Software synthesis offers examples in which the concept of "sound" becomes slippery at best, if not entirely obsolete. The often quoted beginning passage of Jean Claude Risset's "Mutations I" [Risset, 1980] involves the transformation of a chord into a timbre: individual sounds dissolve into one sonority. One could also imagine the case of partials starting long after or ending well before the rest of the other partials which belong to the same sound; a dynamic spectrum is thus created whose beginning in time might not have much in common with its end. Another case would be one of "morphing," where partials of two sounds change allegiance and switch characteristics (or modifiers) from one sound to another.

We conclude that the notion of sound, however crucial to traditional music, has no universal meaning and is not a necessary part of the description of musical events. On the other hand, a complex wave is a universal concept.

4 Implications for music composition

In traditional settings, artists inherit certain ways of thinking which are supposed to help them in their own endeavors. While it is true that nobody can make a clean break with the past, an exaggerated reverence for history may end up inhibiting creative thinking.

A formalism such as the one attempted here introduces a perspective free of stylistic constraints. It challenges the composer's ways of thinking and scrutinizes choices made out of force of habit. A rigorous approach may help understand what one does and why. It may also highlight unexplored possibilities.

rigorous approach may help understand what one does and why. It may also highlight unexplored possibilities.

Considering musical events as complex waves facilitates the view of composition as a homogeneous process that includes similar operations both at the level of sound synthesis and at the level of the entire piece. It also suggests that, at least in the case of electro-acoustic music, the main building blocks are waves (partials), not sounds.

made of only one sound whose fundamental is its total duration of the piece. ”Time intervals” of increasing ”magnitude,” to paraphrase Stockhausen, succeed the three regions discussed previously: cells, motives, phrases, themes, sections, movements (to use the old terminology) represent longer and longer durations and a coarser and coarser granularity. Eventually, the entire piece is one complex wave extending in time.

Seen from the higher vantage point of abstraction, the entire process of composition, including sound synthesis, could be described as comprising: a) a set of elements, and b) an operation which ”associates” these elements in more complex objects. This process repeats itself on various time scales, producing objects at increasingly higher levels of abstraction: partials congregate into sounds, sounds into chords and/or melodies, melodies can enter into contrapuntal relations such as imitations or they may create heterophony, etc. Modifiers such as AM and FM act at the level of a partial, while motive transformations (augmentation, distortion, diminution) operate at a higher level. Someone once said that living beings are made out of a bunch of cells who decided to cooperate among themselves; in music, the composer decides on the mode of cooperation between entities.

We have been trained to think of such entities as defined by pitch. But, why not consider entities (sets) defined on other bases of the vector space and operations acting on time scales other than our Region I, the audio range ? The complex wave formula does not prevent such nontraditional thinking. Nor does it forbid associations that are nonsequential and bridge large intervals of time: some modifiers may create

does it forbid associations that are nonsequential and bridge large intervals of time: some modifiers may create a ”timbre” at the begining of the piece which relates to the product of other modifiers acting towards the end of the same piece, and both determine a third group of modifiers somewhere else, in another time region of the complex wave a.k.a. the piece.

Ultimately, the wave formula and the vector space point toward the possibility of creating non-narrative constructs that are free from the ”words” (cells, motives, etc) and phrases made out of pitches. Music could be more than an analogy for the verbal discourse.

References

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