COMPOSING WITH PARAMETERS FOR SYNTHETIC INSTRUMENTS

BY

CAMILLE MARTIN ANTHONY GOUDESEUNE

B.Math., University of Waterloo, 1989
M.Mus., University of Illinois, 1994

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Musical Arts
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2001

Urbana, Illinois

Abstract

This dissertation presents a theory for designing musical instruments which produce sound by means of computer software, and suggests fruitful ways of composing for such instruments. Unlike other research in new instruments, it focuses on the interfaces between performer, instrument-controller and sound synthesizer. Instead of asking only what sounds a given instrument might produce, it asks how one might perform a given family of software-synthesized sounds, thereby encouraging composition of live music with the rich sound material offered by software synthesis.

After surveying historical practice in explicit composing with parameters and analyzing the parametric behavior of existing orchestral and electronic instruments, a general solution to the problem of controlling an *n*-parameter synthesizer with fewer than *n* (two or three) controls is presented. This solution, simplicial interpolation, is compared to previous partial solutions in the literature. Examples and guidelines are given for applying this solution to particular problems of musical instrument control; future extensions are also presented.

This theory is then applied to design a violin-based software-synthesis instrument called the eviolin. Solo and chamber works using this instrument are analyzed, with emphasis on their explicit representation of parameters in compositional decisions.

Acknowledgments

I thank Herbert Edelsbrunner, Herbert Kellman, and George Francis for braving mysterious foreign terminology; Erik Lund, for patience with my digressions through the years; and Guy Garnett, for asking difficult questions. Kelly Fitz, Lippold Haken, Michael Hamman, Chad Peiper, Eric Rynes, Peter Sleggs, and my colleagues at the National Center for Supercomputing Applications (particularly the Audio Group) shared valuable technical advice with me over the years. Len Altilia, Robert Barry, and Brian Paulson steadfastly encouraged me to complete. Finally, Audrey Fisher, David Schilling, Pappie, Michel, and particularly Mammie quietly kept reminding me—by actions more than words—how insignificant knowledge is when compared to love.

Table of Contents

1. Parametric Descriptions of Music

This chapter presents ways, both theoretical and practical, in which composers have thought of music in terms of individual, separable parameters varying with respect to time. It also considers how effective composers have found these ways to be. The problem is of interest because it directly addresses how to compose for synthetic instruments, that is, musical instruments with a software component.

Minimalism sought in a sense to restart music history. This is revealed by a comparison of the early and later works of Philip Glass and Steve Reich: it is hard to imagine a continuous route linking the standard repertoire to their later works like *Monsters of Grace* and *The Desert Music*. In the 1950's total serialism even more explicitly sought a fresh start. Pierre Boulez in particular wanted to break cleanly with the historical tradition; even Anton Webern's dodecaphony was for him too infected by traditional musical thought.

With the advent of computers as tools for composition and sound generation, a similar break with the past was called for. Of course computers could incrementally extend existing techniques: new technologies often begin by playing the role of older ones and are described in the older language, such as "horseless carriage". But just as *Monsters of Grace* could not have been composed by stepwise extension of previous music, the possibilities of composition and performance with computers were more easily discovered by starting from entirely different points and then (perhaps) incorporating wisdom from the past.

Like particle physicists we can shatter a musical event into its components (half note, middle C, trumpet, *mezzo-forte*); but only as com-posers (putters-together) do we see how these elementary particles might be otherwise reassembled. This atom-smashing analogy misleads a little. Actually, when composers began to use computers they found that computers could deal only with pre-smashed atoms. This comes as no surprise, of course: common music notation grew out of tonal music, became somewhat awkward for dodecaphony, and was commonly abandoned in the 1960's (notably in the scores of John Cage and Morton Feldman).[1] Bringing computers into this realm let music be arbitrary collections of sounds arbitrarily organized, so very different notations or representations of the musical score were needed.

[1] When all twelve pitch classes have equal importance, sharps and flats are notationally cumbersome. Structure would be more clearly communicated by an accidental-free division of the octave (though few performers would begin their musical training with such a notation).

The musical score has become a practical problem again, not merely a notation to supplement the composer's memory, with the more recent use of computers to synthesize sound in real time. When one second of sound can be computed in less than one second of time (instead of in days or weeks!), it can be changed while it is being made: this is exactly a shift of responsibility from composer to performer. As an example of such a real-time system, I have constructed a musical instrument based on an electric violin. The sum of electric violin, computer, loudspeakers, and connective software I call the *eviolin*.[2] Experiments and compositions using it are peppered throughout this dissertation. The eviolin has proven useful in isolating aspects of input gestures, output sounds, and the software-implemented mappings connecting them. Subsequent chapters investigate these three aspects for existing musical instruments, build up a formal language for describing instruments, and rigorously consider the relatively unaddressed problem of mapping. The final chapter unites all these themes, mathematical, instrumental, and compositional, by analyzing the chamber composition *COALS* for eviolin and five other instruments.

We are left with the shattered musical event, and need to reassemble its elementary parameters. Just as with the earliest minimalist music, the strict form of total serialism soon exhausted its possibilities. Parameters could not be composed with in isolation.[3] This leads us to the question: how can we construct *relationships* between parameters (in order to compose with them)? A parallel question is: how can we construct musical instruments (in order to compose for them)? The answers to these two questions, incomplete as they must be, nevertheless have much in common.

1.1 The aesthetics of composing with parameters

Why would a composer in the early twenty-first century want to compose explicitly with parameters? Or why would someone want to listen to or study music in this way? Are there not enough techniques of composing already—serialism, chance operations, minimalism, various kinds of tonality, electronics? In answer, composing with parameters is not an alternative, new and improved, version 2.0 technique. Rather, conscious consideration of parameters can be combined with other ways of composing and analyzing. The sounds in a composition are heard in terms of their varying parameters, and also in terms of poetic associations, allusions to other sounds in the same or another composition, elucidation and development of harmonic, melodic, and rhythmic ideas. But any composition with interesting structure necessarily has sounds which vary. Insights may be gained if, to some extent, this variation is placed in a

---

[2] Changing software arguably defines an essentially different instrument; but I appeal to the analogy of stops and couplers on a pipe organ, which organ is still called a single instrument.

[3] Isolated parameters, ones with few constraints connecting them, result in music of high information density. This leads to the common criticism of total serialism: the information density of its compositions surpasses what can be heard and understood.

formal model—*these* are the things which vary—thus letting us quantify *how much* they vary and also how they vary *together or independently*.

This approach to composition is necessarily incomplete: it is systematic as opposed to intuitive. A piece is generally served well by composing it from both the systematic and the intuitive side. Musical desiderata (or perhaps the constraints of a commission) lead to the choice of a system; constraints imposed by the choice of a system lead to discoveries in the musical surface and ideas for further exploration. These opposing directions tug back and forth. Large swings and reversals of opinion happen in the early stages, and if it appears that a good composition is likely, the swings get smaller as more detailed decisions are made—decisions about the system, or discoveries in the musical surface which the system-so-far has led the composer to.

This particular system (or, rather, way of thinking about systems) is particularly suited to composing with sounds created by means of software, either in so-called batch mode for later audition, or with sounds being controlled and modified in real time. This second possibility has exploded the possible behaviors of musical instruments: perhaps the only generalization about such instruments is that the structure of their software embodies this idea of parameters. The sounds they make, with arbitrary precision of control or range of expression in any nameable or measurable attribute, demand to be composed with idiomatically and not just as extensions of centuries-old idiomatic writing for voice, winds, strings, and keyboards. Their idiom is software, and the common idiom of multilayered software is the function call with a fixed list of arguments—parameters.[4]

Parameters apply not only to the synthesis of sound but perhaps more importantly to its perception. This was not discussed much during the investigation of total serialism in the 1950's, but in the intervening half century the science of psychoacoustics has made much progress. Still, it is clear how much farther there is to go before we have anywhere near a complete model of how humans listen.

1.1.1 Separability of parameters

In a famous article in the popular press, the American serial composer Milton Babbitt argues that contemporary classical music is intended primarily for fellow professionals. He begins with the idea that in such music a single note plays several roles:

---

[4] There are other formal models of software, for example Turing machines, constraint systems, and Petri nets. But the function call is common to everything from assembly language to Smalltalk, and was the dominant software paradigm for the first few decades of computer music.

> “In the simplest possible terms, each such ‘atomic’ event is located in a five-dimensional musical space determined by pitch-class, register, dynamic, duration, and timbre. These five components not only together define the single event, but, in the course of a work, the successive values of each component create an individually coherent structure, frequently in parallel with the corresponding structures created by each of the other components.” (Babbitt 1958, 245)

The emphasis in music has shifted, at least somewhat, from individual notes to “components” or parameters which are presented through the *medium* of notes. (By analogy, critics of representational paintings sometimes regard as merely incidental the figure or object portrayed, preferring to discuss the distribution of light and shade, sharpness of contour, saturation of color, use of perspective, and so on.) Compositions may explore spaces defined by parameters other than the five Babbitt lists here, of course. But the problem only begins with specifying a set of parameters. Fourteen years later, Babbitt (1972a, 7) wonders:

> “…What properties are possessed by an arithmetical progression, for example, which make it appropriate for interpretation as a metrical or ordinal pitch, or durational, or dynamic, or timbral, or density determinant? What data in the field of musical perception determine the rules of correlation? And what of the possible overpredictability and the assumed separability of the components of even the atomic musical event?”

Here he pleads that composers not forget that their elaborate structures eventually result in actual sound.[5] He emphasizes that while we can analyze a single note on the page as a collection of parameters, we intuitively hear it *as* a single note. We see the trees, hear the forest. With practice we can hear the individual parameter streams; certainly performers (and their teachers!) do so as they improve their technique and musicianship. But one would hope that composers write for the holistic listener as well as the analyst. And even the analyst cannot *listen* to all the real-time parameter streams of a composition at once; the parameters not attended to fuse into an organization perceived only subconsciously.

On the other hand we can challenge Babbitt right back: what of the assumed unity of the atomic musical event? Conventional Western notation visually suggests that, say, an eighth-note C# played by one flute *mezzo-forte* is an indivisible unit. But sometimes one marking is not heard as one event: a long note in the context of many short ones; a note marked *sforzando* followed by a crescendo; a sung dipthong whose two vowel colors are contrasted by means of the surrounding musical context. But could one subdivide such counterexamples into smaller and smaller events heard as atomic? Not necessarily:

[5] (Boulez 1976, 64) echoes this concern, in particular about total serial music performed at Darmstadt in 1953-54: “One could sense the disparity between what was written and what was heard: there was no sonic imagination, but simply an accumulation of numerical transcriptions quite devoid of any aesthetic character.” It is not only a question of mapping from one domain to another. The cognitive and perceptual properties of the domain mapped to, sound, must be respected and indeed profited from.

consider the extremely long, unadorned crescendos on one note played by solo clarinet in the movement *Abîme des oiseaux* from Olivier Messiaen's *Quatuor pour le fin de temps*. Such a note admits no clear division into parts, but as it proceeds it nevertheless evokes a number of emotional responses; certainly the note is not heard as a single event. In short, music need not be like chemistry.

But let us be less doctrinaire, less extreme in our examples. In a typical page of Mozart or Babbitt, we hear many things which we would call musical events: chords, phrases, cadences, chromatic or diatonic aggregates, intervals, use of particular registers. A good performance communicates such structure to the audience, sometimes drawing attention to atomic events—this climactic note, that appoggiatura—and at other times dealing with parameters extending over time, bringing out one contrapuntal voice or exaggerating a three-bar crescendo. We really listen in both ways at once, to atoms and to streams of parameters. The individual parameter streams are admittedly "overdetermined," not entirely independent; though this may upset the information theorist it turns out to be useful for the music listener. We now look at some examples from music of earlier periods, and then examine how individual parameter streams interact.

1.1.2 Parametric thought in common-practice music

The core of common music practice in the Classical period was melody, harmony, and rhythm; dynamics, orchestration, and tessitura fleshed out this core, added color to it. True, subtleties of performance practice are critical to even music as restrained as that of Palestrina. But melody, harmony, and rhythm remain central compositionally: these are the elements by which we recognize the structure, the syntax, of a Classical composition.

The *Vivace* second movement of Beethoven's string quartet Op. 135 inverts these priorities. For 47 bars the lower three instruments exactly repeat a five-note figure while the first violin plays an impossibly angular line on an A major triad. Harmony is entirely static, rhythm very nearly so; melody is absent. The structure of this passage is rather found in a long diminuendo from *ff* to *ppp* and a "melody" of tessitura instead of pitch. More generally, this core of melody/harmony/rhythm slowly unraveled as music moved from a vocal model to more idiomatic instrumental models. The greater range and agility of instrumentalists by the time of Brahms let him systematically separate the parameter of register (or timbre) from the parameter of pitch by means of octave displacements, easily found in the last third of his instrumental output. Separation of these two compositional parameters was also regularly seen in the twentieth century.

In the romantic period, the block-like orchestration of early Wagner and especially Bruckner in the manner of separate organ manuals, separate streams of strings, winds, and brass, evidences a conscious choice to use or omit from phrase to phrase. This developed into the fluid, more subtly expressive orchestration of Mahler. By analogy, this suggest that the more general realm of composing with parameters is more interesting when it is explored as a collection of interacting parameters rather than coarsely and individually. A network of streams flowing together and drifting apart, gradually or suddenly, offers richer structure than a field of parallel irrigation ditches. More rigorous evidence for this is presented later on.

Harmonic rhythm offers another example of parametric thinking. Counting the notes of some Classical sonatas reveals more notes per second on average in a slow movement than in a fast movement. From this we see that slow and fast indicate not so much notes per second as chords per second or bass notes per second. On the large-grained scale of movements in a multi-movement work, Classical composers clearly distinguished these parameters of things-per-second. Not all "things" become faster and slower simultaneously in their music.

Of course, Haydn, Mozart and Beethoven would be as bemused by these explanations as they would by nineteenth-century codifications of sonata form. This modern language is ill suited to frankly ancient music; I only wish to show that it also applies to music distant from the worlds of total serialism and synthetic instruments (less forcefully with distance, to be sure).

1.1.3 Total serialism

Pierre Boulez is the most familiar exponent of the school of total serialism, sometimes called integral serialism.[6] Stacey (1987) demonstrates that the aesthetic slant of Boulez in the early 1950's had much in common with musicians, painters, poets and playwrights concerned with breaking with the past. Boulez admits being influenced by individuals such as the following. The early abstract painters Paul Klee and Wassily Kandinsky were the first to truly break with representation; the parallel here is Boulez's rejection of all tonal structures, rhythmic, melodic, and formal as well as harmonic, a completion of Webern's scrupulous avoidance of tonal harmony (particularly triads and octaves) in using pitch class series. Boulez studied with Messiaen; Messiaen and Stravinsky were the major European composers to explore new "nontonal" concepts of rhythm. Piet Mondrian and the painters of *De Stijl* formulated laws of aesthetics called neo-plasticism which were based on oppositions, rather like composing with parameters in the medium of paint instead of sound: vertical/horizontal, color/black-and-white, empty space/volume. Finally, the density and violence in the poetry of René Char and Antonin Artaud made a deep impression on Boulez: he saw in them a transcending of conventional language, in Char's compressed images and Artaud's obliterated grammar and form exposing the subconscious. (Mallarmé and e. e. cummings belong to a later period of Boulez's development.)

In retrospect, Boulez points out the transitional nature of total serialism:

> "This period was fairly short-lived as far as I was concerned, ...this kind of theoretical asceticism and the tough, sometimes arduous work involved brought with it many discoveries; only now, the more I go on, the more I feel that technical considerations are subordinate to the expression I want to convey, and not just a search for method. But at that period there was uncertainty about the evolution of music, and it was not possible to do otherwise. Musical methodology had to be questioned..." (Boulez 1976, 60)

Perhaps the best known composition using total serialism is Book Ia from his two-piano work *Structures*:

> "The first piece was written very rapidly, in a single night, because I wanted to use the potential of a given material to find out how far automatism in musical relationships would go, with individual invention appearing only in some really very simple forms of disposition—in the matter of densities for example. For this purpose I borrowed material from Messiaen's *Mode de Valeurs et d'Intensités*; thus I had material that I had not invented and for whose invention I deliberately rejected all responsibility in order to see just how far it was possible to go. I wrote down all the transpositions, as though it were a mechanical object which moved in every direction, and limited my role to the selection of registers—but even those were completely undifferentiated." (Boulez 1976, 55)

Boulez used this "automatism," this hyper-pure essence of serialism, to purge the past. Much as early dodecaphony could not tolerate accidental tonality in the form of triads, an aesthetic goal of total serialism was a break with past organization of not only pitch but rhythm, timbre, register, durations—any aspect of music which could be formalized and subjected to serial technique, should be.

These compositions were too extreme, too avant-garde to be developed further in the same direction. Boulez used the phrase "at the threshold of fertile lands" in this context (this is in fact the title of the article (Boulez 1955), cited several times in this dissertation). One cannot directly trace the influence of these total serial compositions on others as with bar-to-bar and theme-to-theme comparisons of Beethoven's influence on Schubert; but the idea of splitting Babbitt's musical atom into its components has

---

[6] Sadly, the English term "series" is too entrenched to be displaced: mathematically it is a sequence, not a series (a sum of terms). I avoid the term "set" because that connotes a lack of ordering. "Row" for the German *Reihe* may be even better, but we speak of serial technique not rowish technique.

profoundly influenced composers, particularly those who use computers to help organize their musical ideas.

At least now we know how far we can travel. Boulez has done us the service of exploring these lands, and we (as well as he) can now choose more moderate places along the way. Boulez also agrees with general critical opinion about the paradox presented by the output of total serial procedures:

> "At that point disorder is equivalent to an excess of order and an excess of order reverts to disorder. The general theme of this piece [*Structures*] is really the ambiguity of a surfeit of order being equivalent to disorder." (Boulez 1976, 57)

The equivalence spoken of here is between composition and perception: excessive order in the compositional task produces inadequate order in the perceptual task.

This period of experimentation ended for Boulez within a few years as he grew dissatisfied with the inflexibility and reduced possibility of spontaneous invention within strict total serialism (Boulez 1976, 63). He then polemicized against oversystematization, "Composition and organization cannot be confused" (Boulez 1954).[7] This article was accompanied at the same time by the publication of *Le marteau sans maître*. Griffiths (1978, 36) writes about *Le marteau sans maître*: "Rapidity of change, which is not always a function of tempo, is indeed an important variable in *Le marteau*, and it is in terms of this feature, as well as the cross-references and underlying intervallic constancies... that the work is best understood." He praises "the completeness with which the delirium of a violent surrealism is considered and organized, ... a rational technique's straining to encompass the extremes of the irrational" (Griffiths 1978, 37). This is just like Boulez's admitted paradox of excessive order being indistinguishable from disorder. Griffiths continues: "Its importance lies also in Boulez's discovery, through his proliferating serial method, of the means to create music which neither apes the quasi-narrative forms of tonality nor contents itself with simple symmetries in the manner of *Structures*." Boulez himself writes about the work: "...starting from this strict control and the work's overall discipline there is also room for what I call *local indiscipline*: at the overall level there is discipline and control, at the local level there is an element of indiscipline—a freedom to choose, to decide and to reject" (Boulez 1976, 66).

The "new complexity" composer Brian Ferneyhough also opposes/fuses strict control and intuition. Somewhat like Boulez, he has based his compositional models on poetic texts and paintings (though Boulez tends to obscure rather than reveal the structure of a poem). Ferneyhough realizes this opposition of discipline and indiscipline through prescriptive versus descriptive notations: note heads versus imprecise Italian adjectives, “do this” versus “make it sound like this” (Feller 1994). This notation becomes an encoding which the performer must decode with significant study. One difficulty of this notation is its division of the performer’s attention between multiple levels like tonguing and fingering in separate nonsynchronized rhythms. Beyond this, in some places it follows performance conventions and in other places it opposes them, deliberately contradicting the performer’s extensively trained “natural” way of playing by altering or reversing one parameter of a learned compound gesture. The performer has to suppress old habits of playing and invent new ones: the space between the old and the new habits will thereby “establish audible criteria of meaningful inexactitude” (Feller 1994). Boulez’s indiscipline is manifest within the performer’s gestures as well as in the nonprescriptive notations in the score.

Listeners are extensively trained just as performers are. Pierre Schaeffer, one of the first composers to work with recorded sound, writes that one of the marvels of *musique concrète* was deliberately reversing correlations previously taken for granted, like how the spectrum of a sound would change with loudness (Schaeffer 1966, 539). This breaking down of sound into its parameters (splitting Babbitt’s atom, again) and recombining these parameters differently opened up a vast world of possibilities for composers. With instruments like the eviolin, performers as well as studio composers can play with such correlations and decorrelations.

1.1.4 Scientific formalism

Serial technique was first conceived of as an alternative to tonality for the organization of pitch. In the early 1950’s several composers extended it to organize rhythm, dynamics, register, timbre, modes of articulation—in short, any quantifiable aspect of music. A composition, in order to be taken seriously by other composers, needed a strong contemporary theoretical foundation.[8] By serial or other means, (numerical) organization and systematization were highly valued. This is abundantly clear from leafing through the tables, diagrams and technical prose of articles by these composers in the music theory journal *Die Reihe* (Boulez 1955; Stockhausen 1955; Stockhausen 1957; Ligeti 1958). Across the Atlantic, another paper eloquently defended this idea of composer as scientist, demonstrating that contemporary classical composition was research intended (directly, at least) only for those with sufficient background (Babbitt 1958).

---

[7] In contrast, Babbitt (1972b) defends organization: “A composer who asserts something such as: ‘I don’t compose by system, but by ear’ thereby convicts himself of ... equating ignorance with freedom, that is, by equating ignorance of the constraints under which he creates with freedom from constraints.”

[8] This is an odd reversal. Usually theory lags behind composition but here it leads, the composers acting themselves as theorists.

An aesthetic of science offers useful concepts for the composer: tested, proven hypotheses about density of information, correlation of parameters, statistical reduction of a mass of data into a few numbers, and so on. These powerful results replace the folk wisdom of earlier musical eras. But do they do so completely? Scientific methods were historically developed for scientists, require a scientific language, and flourish in that idiom. Something may be lost in translating them into the language of artists. More accurately, something may be lost in translating artistic ideas into terms usable by scientific methods, to produce scientifically accurate results. Boulez (1954) soon intuited this when he railed against oversystematization and confirmed this with the pliable, unanalyzable language of *Le marteau sans maître*. A decade later Boulez reemphasized the inappropriateness of scientific objectivity in musical composition:

> "Schematization, quite simply, takes the place of invention; imagination—an auxiliary—limits itself to giving birth to a complex mechanism which takes care of engendering microscopic and macroscopic structures until, in the absence of any further possible combinations, the piece comes to an end. Admirable security and a strong signal of alarm! As for the imagination, it is careful not to intervene after things are under way: it would disturb what is absolute in the development process, introducing human error... ." (Boulez 1964, 43)[9]

Systematic thought is useful, indeed necessary, in composing music; but used carelessly it can shield the composer from music (historical music, intentionally; his own, unintentionally).

1.1.5 Electronic synthesis and the saturated listener

In the early 1950's composers suddenly increased their activity of searching for new formal structures and organizing principles. They felt a need for this as a result of having discovered the immense, daunting world of electronically synthesized sounds, in particular the extreme range and extreme precision possible therein.[10] Extreme is exactly right: using almost any measurement possible of sound, electronically synthesized sounds not only extend beyond the smallest and largest limits of human perception, but also vary with more subtlety and fine resolution than the human ear can discern.

---

[9] "...in the absence of any further possible combinations, the piece comes to an end." This seems to directly oppose Babbitt's principle of maximum variety from a minimum of materials, often realized by building sections of a composition to correspond to an exhaustive presentation of a certain kind of object (partitions of a set, for example). But taken in the context of the whole quotation, this opposition holds only if no "imagination" is allowed to disturb the material.

[10] Cause and effect are historically commingled here: some composers at the Cologne studios were attracted to electronic synthesis because it could realize their already formulated organizational ideas.

Babbitt (1962) already made the point that electronics is so liberating that the limits of music are now found in the listener, not the performer. Babbitt (1972a, 9) restated it emphatically (my italics and spacing):

> "Present-day electronic media for the total production of sound,
> providing precise measurability and specifiability of
> frequency, intensity, spectrum, envelope, duration, and mode of succession,
> remove completely those musical limits
> imposed by the physical limitations
> of the performer and conventional musical instruments.
> *The region of limitation is now located entirely in*
> *the human perceptual and conceptual apparatus*... .
> Beyond the *atomic* and easily determinable physical limits
> of hearing and aural differentiation,
> the establishing of '*molecular*' limits on
> the memorative capacity,
> the discrimination between aural quantizations
> imposed by *physiological* conditions
> and those imposed merely by the *conditioning* of prior perceptions,
> and the formulation of laws in the auditory domain
> accurately reflecting
> the complex *interrelationships of the incompatible variables*
> representing the components of the acoustic domain,
> will be necessary before a statement of
> even the most general limits of musical perception will be available."

These "molecular" limits are still very much an open question. Psychoacousticians have made some inroads in the intervening decades, but it may be impossible to even define these variables never mind analyze their interrelationships.

Babbitt's claim was not entirely on the mark in 1972: sometimes the "limitations of the performer," not the listener, still applied. But nowadays such a situation can always be avoided. Sufficient electronic assistance can make the performer as powerful as required, from the simple case of amplification, through instruments like the eviolin, to the extreme case of prerecorded tape.

This state of affairs suggests that the aesthetic of more, the aesthetic of *saturating* the faculties of the listener, has lost meaning. A louder sound, a more virtuosic flurry of notes, a wider range of sounds, a more precise control of timbral details—in themselves these now fail to make an impact, because the listener is saturated. It is in backing off, in deliberately restraining how much of anything (variety, subtletly, information density) is thrown at the listener, that music can happen. The extreme reaction to this saturation, of course, was minimalism. Another reaction comes close to parody: Robert Ashley's painfully loud *Wolfman*. But, like Stefan Wolpe proposes, more moderate directions exist. Wolpe's "unhinging" of dodecaphony is one approach; another is to balance saturation with anti-saturation, drastically restricting the range of some attributes of a composition. Mead (1994) repeatedly illustrates Babbitt's compositional principle of maximum variety from minimum materials, systematically exhausting all possibilities from a single object according to some formalism or other: all the partitions of a number, all the subsets of a set, all the pitch classes of the aggregate. Such a balance is of course already found in the tight counterpoint of J. S. Bach and indeed Josquin des Prez.[11] Growing a large structure from a small germ is also evident in many recent works of *musique concrète* which derive all their material from a small amount of recorded sound (often speech). Berio's *Visage* (1961) is a good example of this extension of the human voice through tape operations.[12]

The indiscriminate use of both electronic engineering and compositional engineering—extreme sounds, extreme structures—can create music which exceeds the molecular limits of our hearing, music properly appreciated only by laboratory instruments and theorists. Nevertheless, engineering has its place in music just as it does in architecture. Many things happen at once as time passes in a composition. A formalism which embraces this advances the bounds of musical possibility, by letting composers work with this truth consciously as well as intuitively. Boulez's threshold of fertile lands may turn out to be not only the limit of where familiar serial grains can be grown, but also the beginning of a soil hospitable to entirely new crops. In traditional twelve-tone composition

> "one could either arrange vertically, in which case one lost control of the horizontal; or one composed with linear series, in which case the vertical simultaneities were very difficult to manage; or else one tried to find a common way to direct both lines and simultaneities, but had to make concessions both ways. The art of dodecaphonic writing lay precisely in managing the dimensions' balancing-act in the confined space... [later, in Boulez's *Structures Ia*] neither dimension is really available any more. But there is nothing terrifying in this, it only means that the possibilities for composition have moved into new territory" (Ligeti 1958, 53).

1.2 Historical overview of composing with parameters

This overview examines particularly the leaders of each school of 20th-century composition which deals explicitly with parameters. Pierre Boulez and Karlheinz Stockhausen epitomize the European approach to total serialism, emphasizing discrete and continuous approaches respectively. In America, Milton Babbitt's compositions and theory extend Arnold Schoenberg's twelve-tone organization of pitch. Iannis Xenakis's equally formidable output of words and music on the application of stochastic tech-

[11] With Bach, this balance might even be similarly ascribed to a conservative reaction against saturation, in his case flashy Vivaldi or Rococo style.

[12] These operations performed on a sounds recorded on tape—change of speed, change of amplitude, reversal—are loosely comparable to the operations of inversion, transposition, *etc*. performed on a twelve-tone series.

niques to composition rounds out the heavyweights. (Compared to these four, Stravinsky's contribution in this area is relatively minor.)

By 1956 when Stockhausen composed *Gesang der Jünglinge*, electronics opened up a vast choice, range, and precision of timbral materials, synchronization, pitch, and spatial placement of sound. This choice of sounds was hardly imagined before then (excluding two famous quotes by Cage and Varèse); the range and precision of sounds would soon surpass the limits of human hearing. This crisis in composition can be compared to that of painting at the dawn of photography: a sudden burst of scientific thought and technology, expensive at first but nevertheless obtainable, shaking previously unquestioned axioms, revealing wide possibilities but with no guideposts as to what would and would not work.

1.2.1 Techniques of parametric organization

In the 1970's and 1980's when computers became more accessible, composers such as Jean-Claude Risset and John Melby wrote tape pieces with the family of languages derived from MUSIC 4.[13] Unlike human performers these languages cannot interpret instructions such as *con fuoco* or *poco meno mosso*. So the list of instructions which the composer gives to the computer is an exact description of the sounds produced, no more and no less. This forces the composer to deal with the many individual parameters of each sound such as amplitude, spectral content, duration, envelope parameters, stereo position, frequency, inharmonicity, or index of modulation. A composition consists of a set of instruments and a list of notes to be played by these instruments. Each instrument has a precise definition of what one of its notes is: a set of usually numeric parameters specifying attributes like frequency, amplitude, attack duration, overall duration, and so on. Some instruments have only a few parameters, others dozens or even hundreds.

How did composers write music in such an environment? And how might they do so in the future?

The composer needs to *organize* these parameters, perhaps by generalizing how he organizes pitch. (Serialism of pitch classes led to total serialism in this way.) Certain parameters of sound material may lie outside such organization and be under intuitive or external control. For example, even Boulez's highly organized composition *Structures Ia* has several aspects which are organized nonserially: emphasis of various pitch registers in different sections, restriction of pitch classes to particular octaves, and palindromic presentation of tempos (Griffiths 1978, 22).

[13] For example, MUSIC 4B, MUSIC 4BF, MUSIC 4C, MUSIC 360, MUSIC 5, MUSIC 7, MUSIC 11. Mathews (1969) and Vercoe (1979) describe the languages MUSIC 5 and MUSIC 11 respectively.

But if we consider sound to be separable into component parameters, we still cannot write entirely "free parametric" music like free atonal music was written in the early 1920's. Some structure is needed, if only to contain the explosion of possible combinations of values which the parameters can assume. Boulez (1955, 47) supports this in the context of electronic music by pointing out how traditional boundaries of music have vanished, all the limitations which grew from the finitude of performers and their instruments. It is tempting to retreat to the familiar world of restricted timbral families, tuning systems, and pulsed meters; not that this is wrong, it simply ignores greater musical possibilities.

Constraints on the value of parameters also help the listener comprehend the structure of a composition. This holds particularly when many parameters are used, as the full space of possible simultaneities becomes intractably large. (Restricting a parameter to a set of discrete values from its entire potential continuum does not significantly reduce the size of the space. Thinking of these discrete values as letters of an alphabet, we can still make an enormous number of words with these letters.)

1.2.1.1 Common music notation

One way to write music in a literal environment like MUSIC 4 is by reverting to traditional notation, using a compiler which converts a language, a textual encoding of common music notation, into these individual notes. Such an ostensibly higher-level language represents things like instrumental lines, pitch classes and octaves, tempo, pulse, accelerando, dynamics, crescendo. But this analogy with high-level computer languages is historically backwards. As common coding patterns were discovered by assembly language programmers in the 1950's, they were formalized and encoded in tools called macro assemblers or compilers (this was called automatic programming). This process of increasing the descriptive power of computer languages still continues. But imposing on a fresh low-level musical description like MUSIC 4 the historical constraints of common music notation is like making it easier to program in iambic pentameter. Such a compiler lets novices quickly produce results from a familiar format, but this benefit may mislead them into thinking that the musical possibilities of the computer are merely those of an accurate but tedious and unimaginative conventional performer. This approach discards most of the flexibility which the computer offers.

1.2.1.2 Correlation of parameters; motifs within parameters

A parameter stream (a single value changing with time) can be described in several ways: (i) as a continuously changing value moving between its minimum and maximum; (ii) as a value changing together with, against, or independent of other parameters; and (iii) as a medium in which motifs can be rendered. The first of these we are familiar with as simple ritardando and accelerando. The second generalizes parallel, contrary, and oblique motion from the domain of pitch to that of pitch (or any one parameter) in the context of many parameters; this is nothing other than what statisticians call correlation and decorrelation. The third similarly generalizes pitch-motifs to parameter-motifs, a motif being a "contour," a short memorable sequence of values which is recognizable as a unit. As pitch-motifs are still recognizable when transposed or when their intervals are expanded, so should parameter-motifs survive translation and moderate scaling of their parameter values.

Motifs may also be called *gestures* or *gestural cells*, particularly if they are contours on a continuum ranging from relaxed to intense. When the parameter is restricted to a set of discrete values from the whole continuum, motifs are more easily recognized as individual entities.[14] Schaeffer (1966, 278) alludes to such motifs defined on a continuum and on a discrete scale built on the continuum. A note—an element of a motif—can similarly be generalized to a set of parameter values more or less fixed over a short interval of time.

Using this terminology, counterpoint can be defined as the combination of several parameter streams to satisfy tight constraints. Just as a parameter stream can be described in several ways, in counterpoint a single note may serve several purposes. It may (i) simply be expressive in an individual parameter, especially if it is near an extreme value of that parameter; (ii) exhibit (de)correlations of several parameters; or (iii) belong to a few motifs (of a few parameters) at once.

The duality between parameter streams and motifs is essentially the same one we analyzed before, listening to parameter streams and listening to atomic musical events. In the ensuing discussion we take motifs to include single "atomic" events. A good model for this motivic way of thinking is found in the free atonal music preceding dodecaphony. Little melodic cells (motifs) remain in this music, freed from the tonal frame but nevertheless contrapuntally combined (parameter streams). For example, the opening of Schoenberg's piano piece Op. 23 No. 1 is built entirely from a three-note melodic cell which falls a semitone and then rises three semitones. Perle (1977, 9–19) presents many examples like this, both horizontal motifs and vertical simultaneities, from Op. 23 and the earlier Op. 11 piano pieces. The interest in these works comes from their variety and freedom of expression, the larger contours of register, timbre, and dynamics along which these motifs lie.

Most people compose not with individual notes (pitches and durations) but with combinations of them. Compositional richness is found precisely in the large variety of combinations, what computer

[14] Implicit in the word re-cognized is the assumption that a motif should recur.

scientists call a combinatorial explosion.[15] Generalizing from pitches and durations, it seems better to compose with parameters not one at a time as in Boulez's *Structures Ia* but by combining them, by varying them sometimes as a unit, sometimes individually. Such opposition between agreement and independence is, for *each* parameter, a source of energy for the music. This is much like the opposition between tonic and dominant driving sonata form or the solo/tutti opposition which energizes concerto and concerto grosso forms.

The opposition between redundancy and high density of information is another source of energy.[16] In tonal and early serial music, information is increased when new material is presented. Paradoxically, the reverse happens in a piece of consistently high information density: immediate repetition of even a tiny fragment grabs the attention of the listener. Ligeti (1958, 60) cites bars 17–18 of Boulez's *Structures Ia* as an example of this. A much earlier example is Schoenberg's almost stream-of-consciousness *Erwartung*. Perle (1977, 19–21) describes how anything at all recognizable in *Erwartung* becomes structural—repeated intervals or chords over a short duration define small-scale areas. Repeated notes often function as punctuation (*i.e.*, structural delineation) in this and other early compositions by Schoenberg and Webern, notably the latter's Op. 5 No. 4.

A concise example of correlation of parameters is found in the second movement of Webern's Op. 27 *Piano Variations* (figure 1). Westergaard (1963) analyzes its serial pitch structure and catalogues its other attributes, three dynamic levels (*p*, *f*, *ff*) and five modes of articulation (staccato, slurred, overlapping quarter notes, and *f* accented trichords. The full twelve-tone series is heard less than its component dyads; the compositional structure derives most basically from intervals, and is clarified by the texture provided by the other parameters.

But when we look at these other parameters, we find little pattern in them individually. It is in combinations of parameters—correlations—where patterns occur and the underlying interval structure is made clear. Here are examples. The note-pair A-A is always *p* staccato. Symmetric dyads around the pitch class A are always found at fixed registers (with a few important exceptions). Successive pitch class series always overlap on their first/last notes. The pair of trichords is always taken from pitch classes 6, 7, and 8 of a series. The lower of a pair of trichords always falls on a downbeat. Finally, we should note

[15] Beethoven frequently writes what is better described as repeated figuration than as melody (the opening of the *Waldstein* piano sonata is a familiar example). Stefan Wolpe also is known for how he creatively develops and varies material of no great intrinsic interest. Both of these attest to the relative insignificance of the things combined, when compared to the combinations themselves.

[16] One could say that information density itself is a parameter to be composed with. But its necessary construction out of more basic parameters and its abstraction make it impractical to work with on the same level as other parameters, for instance reiterating motifs of pure information, or establishing correlations between information density and dynamics or pitch class.

that these correlations obey the constraints presented by the instrument: for instance, accented trichords are always *f*, never *p*.

Figure 1. Excerpt from Webern's Op. 27 *Piano Variations*.

Working with this opposition from moment to moment in a piece can take two forms. First, the amount of redundancy can be varied by simply using or avoiding repetition. This uses motifs, which are possibly atomic in length. Second, information density can be varied by correlating or decorrelating pairs of individual parameters. This second form does not require motifs, but adjusts how large the space of musical possibilities is from moment to moment. The opposition here can be recast as one of more or fewer functional relationships among individual parameters, between agreement and independence. This second form may be conveniently implemented by specifying constraints between parameters: the value of one parameter is determined from another by some fixed mathematical expression, $p = f(q)$. By varying these relationships—the mappings (f), which parameters are mapped (p, q), which parameters are dependent (p), which are independent (q)—as the composition unfolds, interesting structures occur. One might even say that the structures are a side effect of the changes in the effective parameter space as constraints come and go.

An analyst of Xenakis's music states the problem of constraints like this:

> "The problem of dependency or independency is crucial. If variables can remain independent, the coupling is always very easy. If they are *dependent*, ... the couplings generally become much more complicated, but definitely *more interesting*! Defining variables in terms and *degrees of interdependencies* is, I think, a work of vital importance in the development of this kind of compositional attitude. It entails moreover the problem of *hierarchy*: which variable do we choose as the leading variable? Which ones as secondary?" (Vriend 1981, 73)

Schaeffer (1966, 43–44) takes a cave man banging on a rock as the prototypical musical instrumentalist, *musique concrète* from his perspective. From this starting point he demonstrates that the very beginnings of music are found in repetition with variation: holding something constant (say, the objects being banged together) and varying something within that constant repetition (say, the force or speed of the banging). We find a more refined use of this principle in Stockhausen's composition *Kontra-punkte*. Here short groups of notes are defined by holding some parameters constant within the group and varying the other parameters. The notes of a group

> "...have to have at least one characteristic in common. A group with only one characteristic in common would have a fairly weak group character. It could be the timbre, it could be the dynamic: let's say for example you have a group of eight notes which are all different in duration, pitch and timbre, but they are all soft. That common characteristic makes them a group. Naturally, if all the characteristics are in common, if all of the notes are loud, high, all played with trumpets, all periodic, all in the same tempo, and all accented, then the group is extremely strong, because the individual character of each of the eight elements is lost." (Stockhausen 1989, 40)

This is a nearly degenerate use of correlation: a set of constants is in perfect correlation within the group, and these constants are strongly decorrelated with a highly varying value outside the group. This use of grouping musical events (points, in Stockhausen's terminology) can easily be generalized: a group is defined, not by having several parameters constant, but by having these parameters varying so they are strongly correlated with each other and strongly decorrelated with other parameters. (These other parameters are free to be used to define other groups, so one point could belong to several groups.)

Constraints between parameters need not be instantaneous. Stockhausen's well-known *Klavierstück XI* consists of nineteen score fragments, played in arbitrary order as the eye of the pianist roves around the page. Each fragment, as drawn on the page, is followed by indications of tempo, dynamics and touch. These performance indications apply to the next fragment played. Let us label the score fragments $s_1, ..., s_{19}$ and the performance indications $p_1, ..., p_{19}$, indicating their order of appearance in a particular performance with superscripts. Then we write Stockhausen's constraint as: s_j^t implies p_j^{t+1}. For example, part of the piece might be

$$..., (p_2^5, s_4^5), (p_4^6, s_{10}^6), (p_{10}^7, s_{19}^7), (p_{19}^8, s_4^8), ...$$

It is a constraint between the parameters s and p, but a constraint separated coarsely in time. For arbitrary parameters p and q, this can be stated: instead of constraining $p = f(q)$, constrain $p(t) = f(q(g(t)))$. Here $g(t)$ is a remapping of the real line (*i.e.*, time), probably monotonic and continuous. In our present discrete example, if we renotate s^t and p^t as $s(t)$ and $p(t)$, then $g(t) = t-1$ and f is the identity function: the renotated constraint is therefore

$$p(t) = f(s(g(t))) = f(s(t-1)) = s(t-1).$$

These discrete constraints are reminiscent of a Markov process applied loosely to two parameters instead of one (we discuss Markov processes later, in connection with indeterministic control of parameters). We shall see that such constraints are also similar to some advanced notations used by Christian Wolff, though Wolff's notations are less systematic in this regard.

1.2.1.3 Abstract organization

If we compose with explicitly specified parameters, care must be taken when organizing them with extramusical models:

> "…What properties are possessed by an arithmetical progression, for example, which make it appropriate for interpretation as a metrical or ordinal pitch, or durational, or dynamic, or timbral, or density determinant? What data in the field of musical perception determine the rules of correlation?" (Babbitt 1972a, 7)

An arbitrary mapping between domains carries no guarantee that it will produce interesting results. If anything, the sheer number of possible mappings suggests that most of them are uninteresting. To choose a mapping and thereby answer Babbitt's questions, it helps if some structural analogy already exists between the abstract object and the musical parameters to be subjected to it. The more directly acoustic the musical parameters are (timbre and pitch, contrasted with say density or the speed of execution of some process), the more important this is.[17]

A simple example is Xenakis's *Pithoprakta* (1956) for string orchestra. It models the motion of individual molecules in a gas, visible in the many details of the excerpt in figure 2. The glissando speed of individual instruments corresponds to the velocity of individual molecules, and obeys a Gaussian distribution. As the theory of gases describes only velocity, not position, the endpoints of each glissando had to be defined by other means: this other means is what keeps *Pithoprakta* from being merely a sonification of data. The sections of Xenakis's later solo cello piece *Nómos α* are individuated by deriving their parameter values from the structure of the group of symmetries of the cube (Delio 1980). (This composition and the success of this technique we later consider at some length.) As a third example, an abstract formal scheme in Stockhausen's *Mikrophonie I* effectively delineates 33 "moments," relatively independent periods of static musical texture. The sequence of these moments comes from choosing triplets from the three sets: similar/different/opposite; supporting/neutral/opposing; increasing/constant/decreasing (Maconie 1976).[18]

An inappropriate mapping is found in *Structures Ia* where Boulez uses an additive duration series, durations from one to twelve thirty-second notes long. Ligeti objects that this duration series is subjected to the same operations as the pitch class series, when its internal structure is so different:

> "What is unorganic is this pointless transplantation of a system; [pitch classes] labelled with numbers, the dematerialised numbers organised into tables; and the tables finally used like a fetish, as a measure for duration-quantities; thus what were originally mere indications of *arrangement* are now used as indications of *value*." (Ligeti 1958, 39–40)

Many composers have thought to build structures analogous to discrete pitch structure in the spaces given by other parameters. This is reasonable because of all musical parameters it is pitch that has historically had the most structure built on it. The most elementary pitch structure is that of octave equivalence, codified by the serialists in the term pitch class which accurately reflects the set-theoretic concept of equivalence class. Following Boulez we call such structures *modular*. The *module* corresponds to the octave: it contains one representative from each equivalence class, where the representatives are adjacent in the ordering of the dimension being modularized.

---

[17] At the other extreme, Herbert Brün advocates that the overall structure of a composition reflect a desired social or political structure.

[18] Moment form arguably applies at the time scale of 2 to 20 seconds to the second movement of my chamber composition *COALS* for eviolin and five other instruments. These moments are the points in the mesh shown in figure 43.

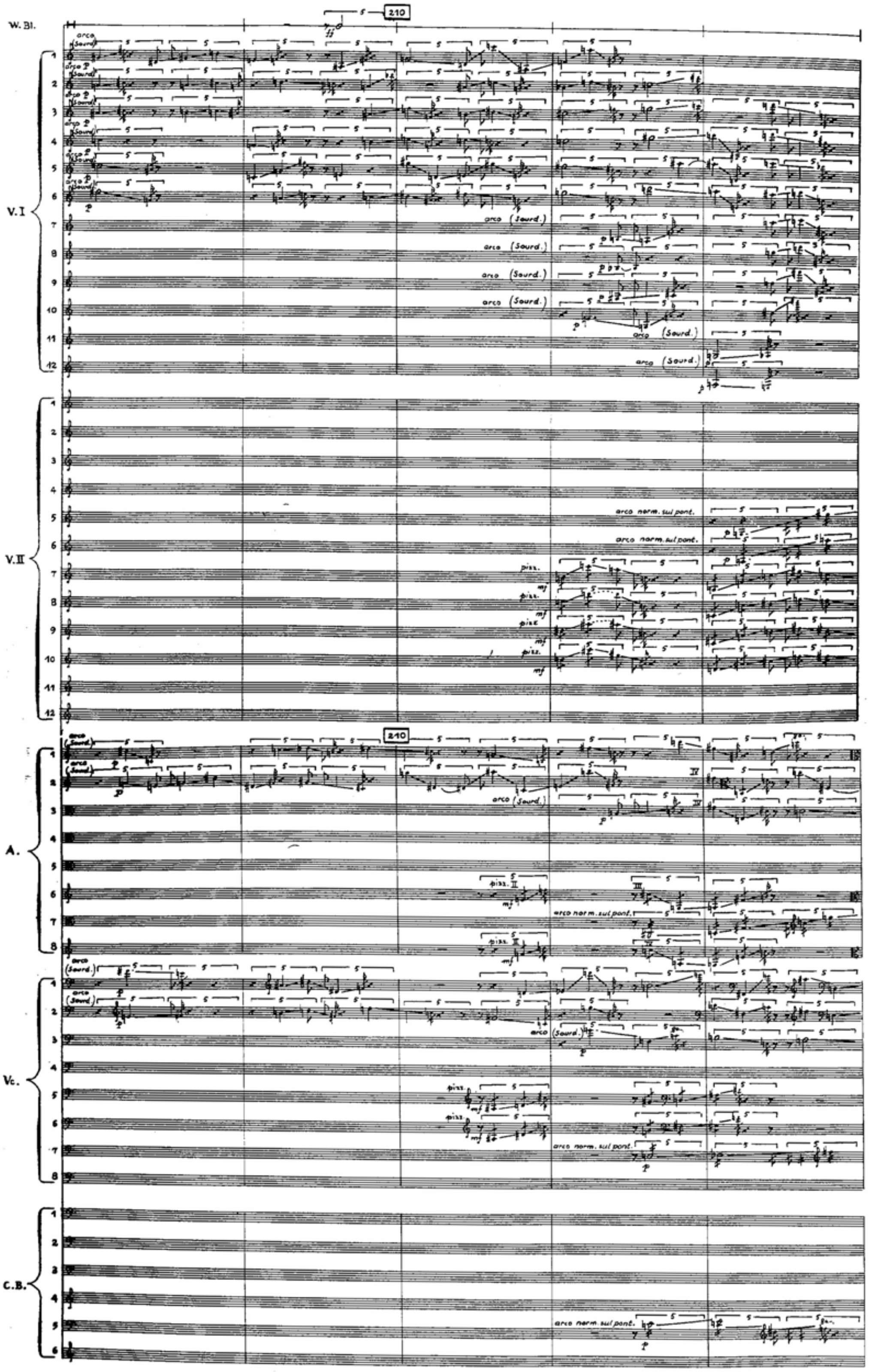


Figure 2. Bars 208–217 from Xenakis's *Pithoprakta*: glissandi representing molecular motion.

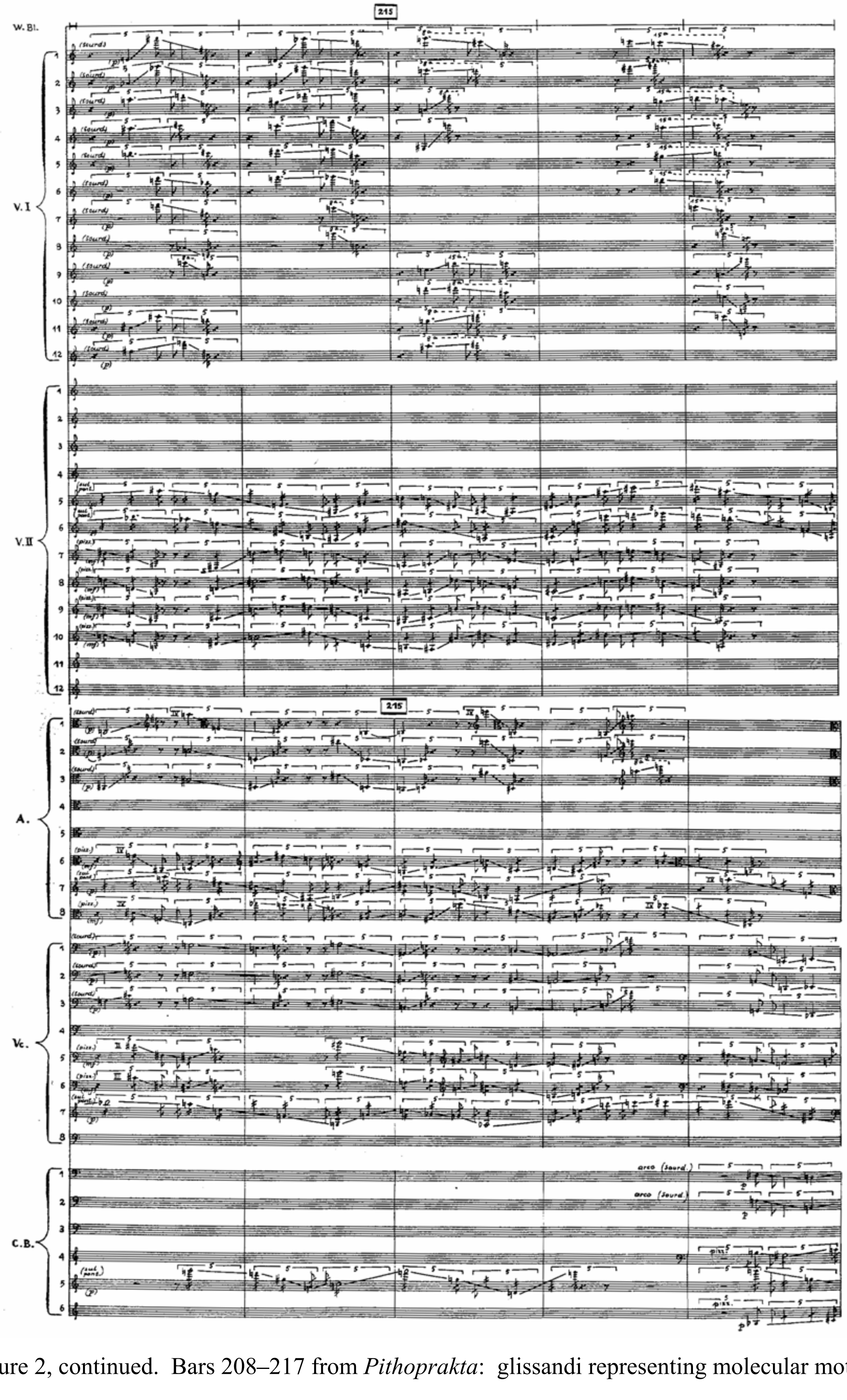


Figure 2, continued. Bars 208–217 from *Pithoprakta*: glissandi representing molecular motion.

1.2.1.4 Geometrical constructions

Another way to manage the composition of music where parameter values must be explicitly specified is by taking the abstract space given by the Cartesian product of the parameter ranges to be literally a geometric space. If we think of a sonic event as a vector of n parameter values, we may equivalently consider it to be a point in an n-dimensional parameter space (a rectangular subset of $\mathbf{R}^n$). Confining the points to certain subsets of the whole space results in a sort of counterpoint of parameter: parameters are not all free to assume all values, but if some parameters are given these values, then some others are forced to have these other values. As one example of such a constraint, my composition *Telescope I* for two-channel tape (1993) investigates a spherical subset.[19] The term "sphere" means the set of all points equidistant from a central point. This spherical restriction is motivated by its avoidance of both boring and extreme sounds, which happens because the sphere divides the entire space into its interior and its exterior. The interior contains sounds with moderate values of all parameters; the exterior contains sounds with several attention-grabbing parameters (which may in fact exceed the limitations of the instrument or of the human ear). So a sound on the sphere itself may have one parameter extreme in value, with others then necessarily moderate; conversely, if most parameters are moderate they force the remaining one to take on an extreme value.[20] This spherical restriction also makes the composer treat the parameters integrally instead of independently, since changing the value of one parameter forces the others to change as well.

What are the inherently mathematical properties of a sphere? We have seen that it bisects the space; furthermore, it induces a reflection between its interior and its exterior (including the point at infinity). Looking at a silver ball, the whole outside universe seems to be contained inside it. The set of rotations of the sphere corresponds naturally to the set of translations (transpositions) in the one-dimensional space of pitch.[21] The rotations of a sphere are complete in a sense: given any two (great-circle) lines of equal length, there exists a rotation mapping one onto the other. This defines parallel (oblique, contrary)

---

[19] This is a 2-sphere in the product space of (spectral brightness, vibrato width, note duration). The parameter values in *Telescope I* are chosen so that each axis has similar noticeability; an extreme value along any of these axes stands out from the overall texture by a similar amount.

[20] A point $(x_1, x_2, \ldots, x_n)$ lies on the unit sphere if $x_1^2 + \ldots + x_n^2 = 1$. So if one $|x_i|$ is large (nearly 1) then the other $|x_i|$ must be small, and vice versa.

[21] These rotations are the rigid isometries of the sphere: distance-preserving, structure-preserving bijections from the sphere to itself. The reflection from the sphere's interior to its exterior is in fact a diffeomorphism (a differentiable isomorphism with differentiable inverse), which can be profited from in extraspherical explorations.

motion from one sonic event to another; a conventional pitch-based motif now corresponds to a path traced out on the sphere.[22]

The boundary of a ball looks locally like a plane, and the boundary of a disc looks like a line. This suggests further structural restrictions like specifying subspheres or linear paths on the original sphere. Related to this, one can project the surface structure of any convex figure in $\mathbf{R}^3$ onto the sphere. For example, a Platonic solid (*e.g.*, the cube) can be placed at the center of a 3-dimensional sphere and project several faces (6 squares) that tile the sphere. These faces, being congruent, have equal area which lends them to stochastic composition. These faces also have the same shape, so by dividing each face into facets (3 or 4 subsquares, for instance) modular construction applies: a face is a module, and a set of corresponding facets is an equivalence class.

The mappings from coordinates in $\mathbf{R}^n$ to sonic parameters should be simple and approximately monotonic, to preserve the sphere's avoidance of both overly moderate and overly extreme sounds. Since the parameters are being addressed holistically, their equal perceptibility is desirable: appropriate choice of range for each parameter ensures this.

A different use of the sphere is shown in Vinko Globokar's orchestral composition *Les Emigres*. Several aspects of this work are structured by means of a rotating dodecahedron, a ball with twelve pentagonal faces. Each face corresponds to a certain kind of music; as the ball rolls, different faces come into view and the amount (angle) visible of each face corresponds to how much of that kind of music is audible. Parameters of this music include register, phrase duration, overall dynamic level, and length of rests in rhythmic cells (Lund 1996). Again, a notable result of this geometrical technique is strongly coupled parameters, a great reduction of arbitrary choice which shrinks an enormous space.

Especially when dealing with motif (or gesture) as a path on the sphere, we distinguish two ways to render the path in sound. In the continuous approach, the musical atom is the motif itself as the parameters of a single sound change, following the path. In the discrete approach, the motif is realized as a collection of short grains which traverse the path. But when working with fundamentally continuous parameters, besides continuous and discrete rendering there is a third possibility which I call *fuzzy discrete* rendering after the field of fuzzy logic (Zadeh 1965; Zadeh 1975; Bezdek 1993). Fuzzy logic deals with precise reasoning about imprecise statements: "John is tall" is certainly true if he is over seven feet, certainly false if he is under five feet, and somewhere between true and false otherwise.

[22] Rotations of the sphere looks superficially like Xenakis's rotations of the cube in *Nómos α* (later in this chapter), but Xenakis's technique is purely formal rather than geometrical: the cube is not embedded in any space of parameters but is used to

A musical example of fuzzy discrete rendering is found in Morton Feldman's notation in works like *Projection II* to play pitches specified only to lie in high, middle, or low register. Another example is given, again, by *Mikrophonie I*. Stockhausen (1989, 76–87) describes his experiments with exciting a large tam-tam with various objects, capturing the sound with a nearby microphone, and then filtering and recording the sound. Amazed at the resulting sounds, he decided to compose for this instrument. He first attempted and then abandoned a detailed score specifying each playing technique. Instead, he chose a set of onomatopoetic words (translating from German to English was interesting and troublesome) and ordered them with respect to high and low frequencies.

Such use of fuzzy discrete rendering is particularly justified by studies of human perception along single dimensions which indicate that people can reliably distinguish, that is, keep in mind, about 5 to 9 different levels (Sheridan and Ferrell 1981, 97–106). Of course, the "just noticeable difference" for almost any perception is much finer than 9 steps over its full range. But if this aspect of the compositional structure concerns not micro-variation but only recognition of which coarse region a value falls into, it is entirely appropriate for Feldman to specify only three "pitches" in *Projection II* even when we can distinguish hundreds.

1.2.2 Pitch organization

First of all, tonality has vanished in music no more than realistic representation has in painting. Like realistic painting it is now one choice among many and is understood as a deliberate choice. Its operation is well described in textbooks; here it suffices to say that its use almost inevitably reminds the listener of older classical or current popular music. Neotonality or "centric" organization loosens the harmonic structures of tonality (no augmented sixth or cadential 6-4 chords). Two common techniques are using alternating static harmonic fields (Stravinsky's 1913 *Le Sacre du Printemps*, for example) and using nontriadic simultaneities as harmonic elements (much of early Bartok). So-called modal music often derives from folk sources, of which Bartok again is the best-known example. Modality here means the use of a set of pitches other than the major or minor scale as the diatonic set. Olivier Messiaen uses modes in this way, but from mathematics instead of from folk music. His *modes of limited transposition* are those pitch-class sets which map onto themselves when transposed by some interval less than an octave. For example, the whole-tone scale can be transposed only by a semitone; the octatonic scale, by one or two semitones. The familiar major and minor "modes" can be transposed by any number of

---

generate permutations and sequences. Also, Xenakis's traversal of a group's multiplication table does not generalize easily to groups of infinite order or other continuous structures like the symmetries of the sphere.

semitones, from 1 to 11, without mapping onto the original pitch set; so composing with one of Messiaen's modes reduces the available harmonic vocabulary. Said another way, using a mode of limited transposition as the diatonic pitch set imposes symmetry and redundancy on the aggregate of twelve pitch classes.

A more famous constraint placed on the aggregate is dodecaphony, composing with twelve tones. As with tonality, the details can be found in textbooks such as (Perle 1977). Arnold Schoenberg's original idea of a series of pitch classes generating other series by means of transposition, inversion, and retrogression—precisely the operations which preserve the interval content of a series while varying its other properties—powerfully changed the role of pitch from establishing, departing and returning to a tonal center, to marking the endpoints of intervals. Babbitt later developed an exhaustive theory of the interval content of series and parts of a series (combinatoriality) by analyzing Schoenberg's music. This systematization let him organize pitch much more tightly in the intricate contrapuntal structures of his own music. Meanwhile, Pierre Boulez extended the serialism of the Second Viennese School in quite different directions from that of Babbitt, developing ways to extend a single pitch class series to a large family of related series. We consider his techniques at length later in this chapter. Stefan Wolpe, less concerned with mathematical rigor than Babbitt and Boulez, emancipated the pitch class from dodecaphony much like Schoenberg was said to have emancipated the dissonance from needing to resolve. Wolpe's dissatisfaction with relentless aggregates led him to compose with series longer or shorter than twelve elements and to "unhinge" the series into parts.

But Wolpe's serialism still uses recognizable pitches. The early composers of electronic music could be described as the emancipators of pitch itself. In *Studie II* Stockhausen constructs individual notes from collections of sine tones spaced, not in a harmonic series or in equal divisions of the octave, but in ratios a multiple of $\sqrt[25]{5}$. This effectively subverts the ability to assign a pitch label to a note. The tape part of his 1960 composition *Kontakte* uses quasipitches, bands of noise that have fuzzy, not-quite-definite pitch. Stockhausen (1989, 110) describes a scale of noise bands in this piece: each band is two octaves wide, and a step in the scale is as large as a perfect fifth in conventional pitches. These two examples link pitch *per se* to the more general notion of frequency content. From our familiarity with sweep filters we quickly suppose that frequency content is a total ordering. But of course this is possible only if we abstract away a sound's frequency content to a single value such as spectral centroid, and thereby ignore much of the sensitive frequency response of the ear. But describing frequency content only in terms of energy per critical band is daunting in how much data is then needed, individual magnitudes for each of about thirty critical bands. An intermediate solution, one which respects the sensitivity of the ear without overwhelming the computer, might describe the overall energy distribution in terms of a few peaks instead of just one (spectral centroid). Each peak can be defined by a bell-shaped curve of amplitude as a function of critical band number, parametrized by an amplitude, mean, and standard deviation; these curves are then chosen so their sum approximates the original energy distribution.

1.2.3 Rhythmic organization

Given a rhythm (a sequence of durations), in his music Messiaen often modifies it with an *added value* or *additive rhythm* by extending one of the notes by a short duration: either extending the note itself, inserting a rest, or inserting a new note. The unmodified rhythm itself is usually absent in the composition. Messiaen also constructs *augmented and diminished rhythms* by simply multiplying each duration by a common scalar; sometimes a rhythm is immediately followed by its augmented variant. He often uses palindromic rhythms, which he calls *nonretrogradable* (though he does not seem to retrograde retrogradable ones!); a concatenation of different palindromic rhythms is characteristic. He also uses *rhythmic pedals*, unchanging rhythms repeated many times. The result of all these is a smoothly flowing line, with somehow predictable attack points despite its remoteness from regular pulse (Messiaen 1942).

Serial composer John Melby has constructed series of rhythmic proportions to which the standard serial operations apply. A rhythmic proportion x:y is read as “x in the time of y beats” where 1 is the quarter note, so 2:3 means two dotted quarter notes and 3:2 means three triplet quarter notes. This is best explained with an example (Dodge and Jerse 1984, 316). Let the original series be (1:5, 5:1, 3:2, 2:1, 1:3). Then we construct its inversion by exchanging numbers in each pair: (5:1, 1:5, 2:3, 1:2, 3:1). Its retrograde is constructed elementwise, (1:3, 2:1, 3:2, 5:1, 1:5). Its inverse retrograde comes from performing both these operations (order does not matter): (3:1, 1:2, 2:3, 1:5, 5:1). Transposition is effected by scaling all durations by a constant factor; we could equivalently say that the reference value of the quarter note is scaled. So the operations of transposition, retrogression and inversion all commute, as is the case with series of pitch classes.

Milton Babbitt’s piano piece *Semi-Simple Variations* works indirectly with durations, concentrating instead on instants of time called *attack points*. This is indirect because the attack points are the endpoints of a duration instead of the duration itself.[23] Mead (1994) shows how each quarter-note beat is

[23] And, curiously, it is the reverse of the movement from tonality to dodecaphony, where emphasis shifted from the end-points of intervals to intervals themselves. Babbitt later refined the technique of this composition into the system of timepoint-ing, as we shall see.

divided into four sixteenth-note attack points, and how Babbitt systematically explores all $2^4 = 16$ subsets of these four attack points.

Babbitt uses another system, *rhythm by partitions*. Each bar is divided into 12 attack points. All the partitions of 12 of a certain size are presented, each partition dividing the bar into, *e.g.*, segments of length 4, 4, 3, and 1 (summing to 12).[24] Each segment is subdivided according to the technique of *duration series*, described immediately below. *Rhythm by conjugate* is related: the bar is partitioned by the conjugate of the partition of 12 used to realize pitch classes, and again each segment of the partition is subdivided with a duration series (Mead 1994).

Just as pitch classes can be concatenated to form a series, so too can durations be concatenated to form a duration series. For example, in *Kontra-Punkte* Stockhausen uses an additive duration series: $\frac{1}{32}$, $\frac{2}{32}$, $\frac{3}{32}$,... to $\frac{12}{32}$ (which equals a dotted quarter note), and a divisive duration series: $\frac{1}{1}$, $\frac{1}{2}$, $\frac{1}{3}$, ... down to $\frac{1}{12}$ of a dotted quarter (which again equals a single thirty-second note).[25] Messiaen and Boulez also use this technique. Duration series of length other than twelve are of course possible, but twelve is convenient for aligning the durations with a pitch-class series. The classical operations of inversion, retrogression and transposition are performed abstractly: having chosen twelve durations and labelled them 0 through 11, the series is then some permutation of (0, 1, ..., 11). We retrogade by reversing the permutation, invert by replacing each element x with $-x$, and transpose by y by replacing each element x with $x+y$, arithmetic being done modulo 12.

An early work of Babbitt's, his second string quartet, uses duration series only four elements long but with the classical operations as defined above. The resultant durations are rendered as attack points, typically with respect to a pulse of sixteenth or thirty-second notes. Mead (1994) notes that transposition of these duration series destroy what we would call their interval structure, were they pitch-class series; this makes transpositions very difficult to recognize as related to the original form. He also notes that only a few forms of the series could be used to form pairs of series related by hexachordal combinatoriality, where the two "hexachords" have the same overall duration.

Babbitt (1955) later rejected the serialization of durations, criticizing independent pitch and rhythm series as simultaneity rather than polyphony, resulting in harmony that was merely fortuitous. He preferred his new technique of *time-pointing* for its clearer analogy (almost isomorphism) with dodecaphony—pitch to time point, interval to duration, octave to modulus or bar-duration:

---

[24] A partition of 12 of size 4, for example, is four numbers which add up to 12. We describe this rigorously later on.

[25] The later piece *Gruppen* arguably extends this idea to metric modulation (Maconie 1976).

> “The recognition of the permutational character of twelve-tone operations suggests their application to orderable non-pitch elements such as duration, dynamics, *etc*. The twelve-tone application by analogy demands, for its feasibility, rules of correlation which assure the empirical interpretability of these invariants which inhere under the pitch application of the operations employed; such rules have been defined.” (Babbitt 1972a, 8)

Time-pointing begins by defining twelve (equally spaced) attack points in a bar. A time point series is again a permutation of (0, 1, ..., 11); but now the numbers 0 to 11 mark attack points, not durations. Successive attack points of the series may be realized within the same bar, or they may have a bar line between them (figure 3). We treat this influential technique at length later in this chapter.

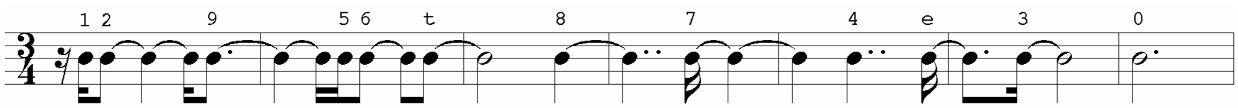


Figure 3. A rendering of a time point series in a modulus with twelve sixteenth-note attack points. (The letters ‘t’ and ‘e’ conventionally represent ten and eleven.)

1.2.4 Timbral organization

The classical treatises on orchestration by Berlioz and Rimsky-Korsakov and the scores to the Mahler symphonies describe (among other things) which instruments blend well or contrast effectively. These texts present the design of timbre as a combinatorial art, and arguably an additive one as the orchestra grew in size from Mozart’s day to that of early Schoenberg.

After the collapse of the orchestra into the chamber ensemble, Pierre Boulez (1963b) offered an example of tight organization of such an ensemble in *Le marteau sans maître*. Its instruments, various though they are, can be connected in pairs: contralto voice and alto flute share human breath, alto flute and bowed viola monody, plucked viola and guitar the plucked-string timbre, guitar and vibraphone long resonance, damped vibraphone and xylophone pitched mallet timbre. This instrumentation also alludes to the widely imitated “Pierrot ensemble” of Schoenberg’s song cycle *Pierrot Lunaire*.

Composers of electronic music in the 1950’s were excited about the possibility of composing timbres directly and hoped to build analogous structures from the level of waveform all the way up to large-scale form. Timbre turned out to be harder to create than expected, though: many practical problems lay between the theory of Helmholtz and actual synthesis of timbres as rich and subtle as what performers expected from their own instruments. Timbral organization was again additive, but the orchestration was one of laboriously controlled sine tones and noise bands (*e.g.*, Stockhausen's compositions *Studie II* and *Kontakte*). Synthesis of "interesting" timbre became practical only when sounds could be organized with the help of software—from the Moog modular synthesizers, through special-purpose digital signal processing hardware, to pure software systems. In the middle of this period of implementation, Grey (1975) introduced to the computer music community the concept of a *timbre space*, a continuous family of reasonably diverse timbres indexed by two or three perceptually defined axes (spectral brightness, synchrony of onset of harmonics, and attributes of the attack such as inharmonicity and amount of noise).

Both Boulez and Stockhausen applied serial and permutational organizations to timbre, often starting with a linear continuum of timbres: timbres could be chosen continuously along this line, or in (perceptually) equally spaced steps analogous to the equal semitones of tempered pitch. This idea was developed into *Klangfarbenreihe*. Boulez (1955, 55) remarks that orchestration had by that time developed from a decorative function to a structural one. Stockhausen's tape piece *Gesang der Jünglinge* is an early example of elaborate timbral structure (additive / constructive, to be sure). In it serial and permutational methods are applied to eleven such linear continuums: vowels, periodic amplitude modulation, stochastically defined pulse trains, and so on (Stockhausen 1955, 57).

Acoustically produced timbres have been organized similarly. A single example, Stockhausen's *Mikrophonie I*, illustrates both continuous and fuzzy discrete specification of timbre. Some notations in the score direct the performers to vary certain timbre-affecting parameters continuously (distance from microphone to an object, intensity of a scraping motion), others discretely (three ranges of frequency content). But the compositional organization of timbre until very recently has not been much more sophisticated than nineteenth-century organization of dynamics or tempo: the status of timbre as the psychoacoustical "junk box" containing everything not otherwise measurable has made this inevitable.

1.2.5 Spatial organization

Historically we observe a trend from antiphonal placement to more sophisticated gradations and movements of sound. Heeding Babbitt's dictum that any recording is a work of electronic music, this is evident when comparing early stereo recordings by the Beatles with precisely controllable software-based panning used in contemporary multitrack production of popular music. Choral music, the first notable composed use of spatial separation (Giovanni Gabrieli), has similarly developed.

Orchestral music stayed for many years at the level of "one horn offstage." Only in the latter half of the twentieth century have multi-orchestral works like Stockhausen's *Gruppen* and *Carré* extended the spatial organization of Berlioz's large works like the *Requiem*. Cross-fading is a notable extension to spatial technique found in these recent works. Iannis Xenakis is the composer best known for his spatial experiments in works such as *Nómos γ* and *Terretektorh* which situate a large orchestra among the audience. *Terretektorh* divides the players into eight sectors of a disc, among which sounds can fly back and forth.

Chamber music depends less on all musicians seeing a single conductor, so it has exploded the most in the spatial realm. Xenakis's 1969 work *Persephassa* for six stationary percussionists is a tour de force of rotating sounds. Many of Vinko Globokar's compositions involve, in more or less deterministic ways, soloists situated around the concert hall or moving through the audience; a recent example is his 1995 *Masse, Macht und Individuum*. An extreme kind of spatial organization, where performing groups are separated so the audience must move from group to group, is given by the performers in different rooms in Stockhausen's *Musik für ein Haus* performance at Darmstadt in 1968 (Stockhausen 1989, 157), or Elliott Schwartz's *Elevator Music* for performers on twelve floors and audience in an elevator, sound "leaking" between the floors (Cope 1971).[26] Such compositions are inherently indeterminate in the freedom of the audience to move around. But they also involve indeterminacy at a deeper level, insofar as they must adapt to whatever buildings are available to avoid of the expense of constructing a special building for each performance. (A celebrated example of such a building, though with only one room, was the spherical auditorium of the German pavilion at the 1970 Osaka Exposition, described in some detail in (Stockhausen 1989, 103). Griffiths (1994, 159) has a photograph; the score to Stockhausen's *Spiral*, Universal Edition 14957, includes several photographs and diagrams.)

Multi-loudspeaker playback became more common in the 1990's, particularly for tape music at computer music conferences. A two-track tape may be realized on such a *sound diffusion* system by a specialist (a performer, really) in collaboration with the composer; multitrack compositions may also be designed for a particular setup. Swooping sounds rotating around the audience, particularly easily localizable pulse trains, have long remained fashionable in tape music. As with other attributes of sound, we can roughly divide composers into continuous and discrete camps, here championed by Xenakis and Boulez respectively.

1.3 Serialism and total serialism

We now consider in detail the approach of Pierre Boulez. This focus is justified because no other composers have written so broadly, systematically, and exhaustively about their discoveries growing out of classical dodecaphony. Certainly much of Boulez's writing is polemic, but to be fair much of Stockhausen's is anecdote. After stripping away what is no longer relevant after some decades, Boulez's corpus of prose remains impressive. Only Milton Babbitt's essays come close in depth as well as breadth.

Boulez splits the musical atom into five elementary components: pitch, duration (and tempo), dynamics, timbre, and location in physical space. Each of these components can be subjected to serial procedures. Each of these can inhabit a space that is either smooth or striated. Together, they can be combined into larger structures. Here we consider in some detail how Boulez and Stockhausen use these components, individually and in combination.

1.3.1 Serialism of parameters

A series is defined as a hierarchy "with a view to organizing a *finite* ensemble of creative possibilities connected by predominant affinities, in relation to a given character; this ensemble of possibilities is deduced from an initial series by a *functional* generative process" (Boulez 1963a, 34 (his italics)). Boulez points out that basic attributes of sound (pitch, duration, dynamics, timbre, spatial position) as well as combinations of these can be serialized.[27] He emphasizes finitude, in our terminology the discretization of a continuum: his thoughts tend more to organizing symbols in an alphabet than to continuous variation. He repeatedly emphasizes his disgust with the continuum: pitch glissandi, filtered white noise, spatial glissandi (sounds swooping and whirling). Pitch clusters he dislikes too as a milder form of filtered white noise, so when he presents a pitch structure with adjacent pitch classes he avoids semitones and instead voices it with minor ninths and major sevenths.

But Boulez strongly supports separation of parameters (these five basic types and derived ones) and their deliberate organization: primary serialization of one and secondary serialization of others; combined serialization of several; and even mere coexistence of some parameters (Boulez 1963a, 35). Of these five basic parameters, pitch and duration are historically primary; this can be seen in the historical development of music notation. Each parameter has its own degree of precision in notation and performance—a single notated mark corresponds to a "field" or range of possible performed values (Boulez 1963a, 37). To discretize the continuum in a way which is useful for serialism, Boulez imposes a modular structure on the individual points by dividing the total range into modules like the piano keyboard is divided into octaves. Presumably the modules are contiguous, have the same size, and

---

[26] A simpler musical effect can be heard at the Thursday lunchtime concerts at the Beckman Institute of the University of Illinois. The musicians playing in a large atrium are heard in the elevators as pianissimo, piano, or forte as the elevator is between floors, at a floor, or has the doors open.

partition the space (include all the points of the space without duplication). These conditions, together with the natural order on the space of individual points, induce an equivalence relation on the space isomorphic to that induced by the modulo-n operator on the set of integers $\{0, ..., kn-1\}$. Time-pointing is a common example of modularized duration. Timbre too can be divided into modules. In my six-choir piece *Melisma* (2000) the module is the voice, soprano, alto, tenor, or bass; each module contains a set of timbres, dark and husky (low range), comfortable, straining, and falsetto. In this structure timbre is a more primary parameter than overall pitch, high or low.

By analogy with the concepts of pitch and pitch class, Boulez defines the "relative value" and "absolute value" of a parameter as respectively its actual value and its value modulo this implicit equivalence relation. He also distinguishes "fixed and mobile density of generation," which possibly means equal versus unequal spacing of points in the continuum, or possibly modules of constant versus varying size. (It is difficult to reconcile the scientific style of his prose with its profusion of terms which are undefined or even contradictory to accepted technical usage. Perhaps he wants his words to be analyzed no more easily than his music, a defense against other composers' blind overuse of systematization.)

Given a simple twelve-tone pitch class series, partitioning it into 3+2+4+2+1 segments yields five new sets of pitch classes. Alternatively, the sequence of intervals from a pitch class series can serialize another parameter like duration or dynamics (Boulez 1963a, 40). Combining parameters in ways like these produces a score which looks fearsome in conventional music notation but simple, even terse, in algorithmic notation. This algorithmic simplicity may not be apparent to the performer or listener. Even the theorist may have difficulty reverse-engineering the compositional method: (Northcott 1971; Griffiths 1978; Stacey 1987) all repeatedly point out the impossibility of thoroughly identifying twelve-tone series and serial operations in Boulez's later music, and indeed *Le marteau sans maître* was first comprehensively analyzed only by Koblyakov (1990). The surface of this music is intriguing, as is the case with Babbitt's later compositions: it is far from tonal in both pitch and rhythm, far from even the intervallic organization of dodecaphony, but still it is somehow perceptible as tightly organized. The compositional challenge is to keep the music from stagnating into a single undifferentiated percept, to keep listeners "inside" the piece instead of aloofly observing a performance which happens to be in the same room as they are.

Though not claiming mathematical rigor, Stockhausen does think in terms of separate parameters which are components of an overall space, which he calls unified musical time. "[T]he different

---

[27] This distinction between primary attributes and their combinations relates to my distinction between inherent and noninherent parameters.

perceptual categories, such as color, harmony and melody, meter and rhythm, dynamics, and ‘form,’ must be regarded as corresponding to the different components of this unified time.” (Stockhausen 1962, 42)

> “[Stockhausen] had expanded and developed the serial principle, finding that he could work with scales of duration, tempo, instrumentation and even, in *Gesang der Jünglinge*, a scale between vocal and synthetic electronic sound. What now emerged was the idea of smooth mediation between extremes in place of the quantum jumps of serialism.” (Griffiths 1994, 142)

This is reasonable: we perceive pitch much more precisely than other acoustic dimensions (here, duration, tempo, and timbre). So correspondingly, the melodic contours (changes of value with respect to time) of these other parameters will vary less from moment to moment than those of pitch. The high level of organization of pitch in dodecaphony is arguably too great for it to be applied directly to acoustic parameters less precisely perceived. “As far as Stockhausen was concerned, the positive lesson of total serialism was that any aspect of sound could be subjected to serial transformations” (Griffiths 1994, 139). In the sense of a weaker serialism, these other parameters could be organized just enough to be appropriate for each.

In *Gesang der Jünglinge* Stockhausen needed a wide range of timbre in both speech and electronics. Organizational techniques in this composition include: series of vowels and of consonants; sine tones presented serially; periodic and stochastic amplitude modulation and frequency modulation; filtered and stochastically modulated noise; filtered pulses; and periodic or stochastic pulse trains (Stockhausen 1955, 57). This powerfully illustrates the explicit construction of sound (in this case, recording tape) from separate parameters.

Stockhausen (1989, 80–81) describes the score of his *Mikrophonie I* for the “exciter who plays on the surface of the tam-tam, ... the microphonist, and ... the player of the filter and potentiometer.” Parametric structure already begins here with a threefold division of labor. Both the exciter and filter control a parameter of register, divided into high, middle, and low. The microphonist controls the parameter of distance between point of excitation and microphone measured parallel to the surface of the tam-tam, similarly divided into three ranges. Stockhausen adopts a time-line notation where lines drawn on the score indicate sound, like a graph of frequency versus time.[28] The line of the exciter gets thicker (not in three gradations, but smoothly—continuously, again) to indicate loudness. The line of the microphonist also varies continuously in thickness to indicate how close the microphone should be to the surface of the tam-tam.

[28] Such notation goes back to his postcompositional published score of *Studie II* for tape, and of course far earlier than that in the physical sciences.

This coarse resolution of only three gradations is also found in some of Morton Feldman's compositions where pitches are specified only as high, medium, or low. In the case of Feldman it is even more obvious that the coarseness is not so much an attribute of the sound as it is an optimization of the communications channel from composer to performer. (Feller (1994) notes that Ferneyhough too considers music notation to be an encoding which the performer decodes. This encoding has prescriptive components like note heads and descriptive components like imprecise Italian adjectives with poetic connotations.) When we see coarse notations like these for a parameter, we can conclude two things. First, the composer trusts the performer(s) to use good judgement in precise choice of value for that parameter. Second, notating the value more precisely would be needlessly fussy or even mislead the performer's search for musical meaning.

1.3.2 Pitch

Boulez elaborates the structure of the pitch series with methods like frequency multiplication (see below) to produce richer pitch structures, and just as parallel fifths and octaves are avoidable in four-part but not eight-part counterpoint, these richer structures will sometimes incorporate octaves and triads. Early twelve-tone music avoided such structures lest it be misunderstood for even a moment as tonal. Boulez's music avoids distracting the ear with unintended tonal meaning by placing such potentially ambiguous structures in contexts like short duration, high density of pitches or events, and varied timbre. This loosening up of classical dodecaphony is taken even farther by Wolpe, who complains (at least in the sphere of pitch) about serialism's characteristic grayness, its maximal information, its maximal depth of structure and communication. As one critic puts it,

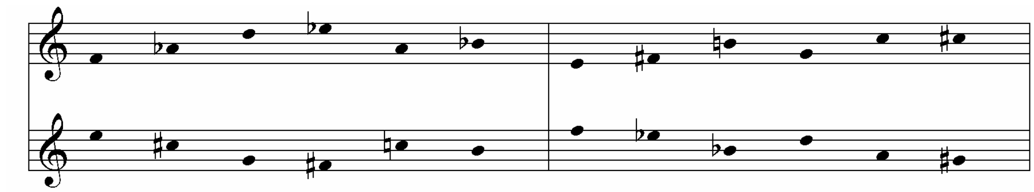
> "[I]t is extremely difficult to eliminate incidental tonal inflections from, at least, simpler serial textures. ...The degree of textural complexity and rapid manipulation of the total chromatic required to ensure freedom from tonal adulteration must prove in the long run intolerably constricting in other ways." (Northcott 1971, 34)

This grayness can be explained less poetically as the natural fusing together of details which are not yet able to be discriminated and assimilated into one's musical experience to date. In particular, rearranging a series of twelve pitch classes by means of transposition and inversion misses the point for Wolpe: "No amount of [serial operations] can relieve the ear from hypertrophic abundance of a pitch-totality that, in this exclusive form, must stagnate" (Wolpe 1959, 277). The harmonic rhythm itself is heard as unchanging, and working within the series cannot change this. So instead, "The all-chromatic chain can be unhinged, its sections interrupted, isolated, and arrested" (Wolpe 1959, 277). Pitch class series longer than twelve elements are possible and indeed appropriate in this world of variable harmonic rhythm. Reduced series from pentatonic pitch sets or Messiaen's modes of limited transposition are also possible.

When we encounter an incomplete pitch set in such a context, either it is a postponed completion of the aggregate or else it stands outside the logic of aggregates unhinged and perhaps later hinged back in. The slower "speed of complementary circulation" (harmonic rhythm, in a sense) more strongly focuses each pitch region. Wolpe's *Passacaglia for Piano* demonstrates one way of working like this: a six-note series, presented successively with classical operations, very gradually completes the aggregate. This work also uses "one-interval" series built up from successive single intervals (as opposed to the all-interval series favored by Babbitt and his followers).

Wolpe partitions the series of pitch classes, like Babbitt or Boulez, but not for building systematic elaborate structures as they do. His operations tend towards intuitive expression (the register, dynamics, rhythm, or character of a pitch class), or changing the identity of a pitch class in the series by changing its context: as part of an interval, repeated, not repeated, isolated out of a chord, always in the same register, varying over all registers, alternating between two registers.

Wolpe often presents a series (even theoretically, outside the context of a complete composition) with extra structure as an inherent part thereof. This structure can take the form of segmentation, dynamic contours, expression (Italian, no, English adjectives), or in "dispositions" of segments or individual notes: in three ways, wide registral spacing, canonic plus vertical. So in the composition, every presentation of a segment comes with that extra baggage. To lend variety to the musical surface Wolpe generally uses these dispositions and segmentations instead of standard operations on the series.



(a)

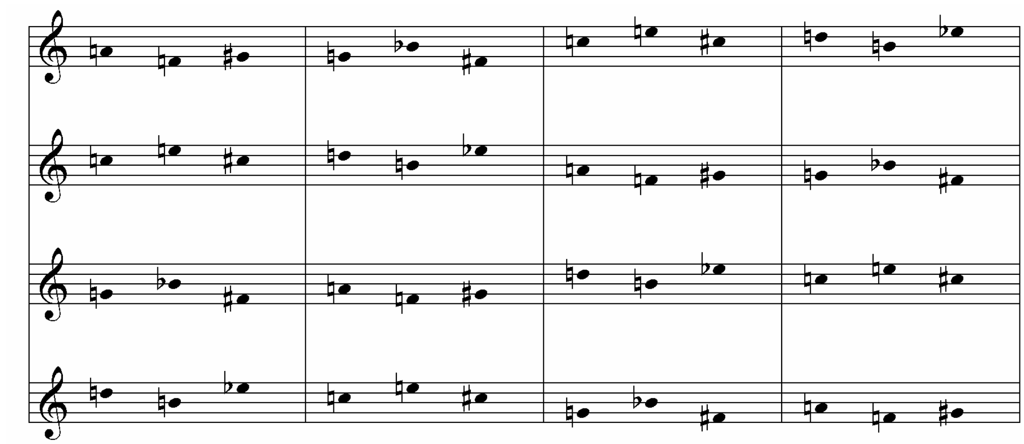
A B C D

C D A B

B A D C

D C B A

(b)



(c)

Figure 4. (a) Two IH-combinatorial series. The lower is an inversion of the upper; the first hexachord of the lower one is the same as the second of the upper one, and vice versa. (b) A block of four horizontal lynes. (c) The same block, rendered as actual pitches.

A far more rigorous method is found in Babbitt's construction of *arrays* for statements of twelve-tone series. A pitch class series presents one aggregate of the twelve pitch classes. Two IH-combina-torial series, presented simultaneously, contain one aggregate each horizontally; but vertically, the pair of first hexachords and the pair of last hexachords also contain an aggregate (figure 4a). Babbitt goes one step farther: a *trichordal array* presents four series simultaneously, each series still presenting an aggregate, but with many more aggregates. Let A, B, C, and D be a trichord, its retrograde, inversion, and retro-grade inversion respectively, such that ABCD is an aggregate. (Such a construction of a series is due to

Webern.) Then in a *block* of four *lynes* (figure 4b) the four columns, the four 2×2 corner squares, and the four 2×2 edge squares, as well as the four rows, are aggregates (figure 4c). Parameters other than pitch class can emphasize particular aggregates or hexachords over others, or simply contrast them. For example, each lyne can be played by a separate instrument (timbre), the four corner squares can be *pp*, *p*, *f*, and *ff*, and the columns can occupy different pitch registers. (This construction is a special case of a Latin square, an $n \times n$ array of symbols $x_1, x_2, ..., x_n$ with the property that each row and column is a permutation of $\{x_1, x_2, ..., x_n\}$ (Denes 1974). From this more general structure we can imagine creating 6×6 dyadic arrays.)

We can concatenate this block with a permuted copy of itself, exchanging A with B and C with D. There are eight partitions of a set of 4 elements of size 1 or 2, so we can apply these partitions to the eight 4-element columns we have just made. This yields what Babbitt calls eight partitions of two blocks of a trichordal array (figure 5).

```
A   ,    B ,    C ,    D ,    B ,  A   ,    D ,   C
  C ,  D   ,    A ,  B   ,  D   ,  C   ,    B ,   A
  B ,    A ,  D   ,  C   ,  A   ,    B ,  C   ,   D
D   ,  C   ,  B   ,    A ,  C   ,    D ,    A ,   B
```

Figure 5. Eight partitions of two blocks of a trichordal array.

Each A,B,C,D is again a trichord and all ABCD’s are aggregates: horizontal, vertical, 2×2 squares, even two diagonals. Such a block can be presented in the musical surface in many ways, for instance by dropping a few pitch classes from one lyne and adding them to another lyne. In this case the columns of the block will be no longer four trichords but rather some other partition of 12 like {3, 3, 1, 5}.

Babbitt often renders lynes of time points with dynamics and lynes of pitch classes with register (Mead 1994). Analogous contours of pitch and duration can then unfold at different rates.

1.3.3 Duration and rhythm

Two kinds of duration appear to be intended in (Boulez 1963a): first, the conventional duration of a sound or note, the interval of time from its onset to its fairly sudden termination; second, the interval of time between successive note attacks. The latter is more appropriate for decaying (percussive) sounds which have no definite end; in a musical context the effective duration of such a sound is measured from its onset to the onset of some later sound.

The familiar way of constructing a variety of durations which we see in common music notation is to start with a fixed quantum (say, a quarter note) and then construct multiples of it (half note, dotted half note) or fractions of it (sixteenth note, triplet eighth note). Again Boulez makes analogies with pitch: the quantum semitone is multiplied to make the intervals of the chromatic scale and divided to make microtones, and tessitura in pitch has a correspondent in duration. (He later points out that such strong analogies fail with dynamics and timbre because they are performed and notated less precisely than pitch and duration. This is of course due to the relative imprecision with which we hear dynamics and timbre.)

A collection of durations may be presented as a series, not necessarily with twelve elements. Boulez distinguishes three ways of modifying such a series: (i) fixed, multiplying all values by a constant (commonly called augmentation or diminution); (ii) mobile, adding or subtracting a constant to all values; (iii) mobile evolutionary, modifying each value according to some nonconstant function of the series itself.

Given a series of durations Boulez proposes four ways of distributing these durations in time to produce a "block" of durations (Boulez 1963a, 56). (i) "Symmetric distributions" are left-, right-, or center-justified with respect to a fixed moment in time. Variations of this are: left-justified followed by its retrograde, leaving a hole in the middle; superimposing several of these; right-justified followed by a "bar line" and thereafter left-justified. (ii) "Symmetric irregular distributions" are as above, but with respect to a diagonal line instead of a vertical bar line. Formally, the point in time about which the symmetry is placed is no longer a single point, but one which varies with respect to some parameter of the sounds placed in these durations—their pitch or their dynamics, for instance. (iii) "Asymmetrical distributions" have no symmetry at all. (iv) Finally, each individual duration can be subdivided into smaller durations, for example replacing a half note with eight sixteenth notes; pockets of silence can be introduced by omitting some of these smaller notes.

Combining multiplication and division of the quantum duration produces a rich structure, Boulez notes; this is precisely the structure of the rational numbers which is lacking in the integers taken together with their reciprocals.[29] All these techniques can be applied recursively in the manner of a context-free grammar: duration blocks of duration blocks, of linearly presented duration series, of fragments (elements of a partition of) a series. In particular, the opposing forces of assembly (blocks of blocks) and disassembly (partitioning) produce rich structures, like multiplication and division produce the rich structure of the rational numbers.

Can durations be given a modular structure? Since there is no quantum of duration, just like there is no quantum of pitch interval, logarithmic time scales can be constructed (Boulez 1955, 51). The module is, like the pitch-octave, a doubling, so durations of 0.5, 1, 2, 4, and 8 seconds correspond to five consecutive C naturals on the piano keyboard. Of course this structure is already evident in Mozart's hierarchical 1-, 2-, and 4-bar phrases, but now we go farther: between each pair of these lengths we define other lengths, related logarithmically. So we have a structure where ratio of durations corresponds to interval between pitches. Unfortunately we may not be able to hear the equivalence of two different members of the same duration class. We recognize pitch octaves because the overtones of two pitches align most closely (the cochlea resonates most simply) when the pitches are in an octave relationship. No psychoacoustical parallel is to be found in rhythms below 20 Hz as with pitches above 20 Hz. Granted, two pulse trains, one with period a small-integer ratio of the other, are heard as nearly related (even if they are out of phase). But if we consider single durations rather than successive repetition of durations, no percept occurs as strongly as the recognition of two C naturals in different octaves. (Stockhausen seems to have intuited this: *Gruppen* is the first of several of his compositions which use a "chromatic scale" of tempos within a 1:2 module (Stockhausen 1989, 41).[30])

Another difficulty with modularized duration is the coarseness with which we perceive duration. If the range of durations is from 0.05 to 60 seconds, and the "semitone" is a ratio of 3:4, we have only $\log_{4/3}(60 \div 0.05) \cong 25$ steps to work with, considerably fewer than the number of pitches we can recognize. A final difficulty is that, unlike simultaneous pitches, simultaneous durations cannot be presented clearly. Only in a strictly hierarchical presentation of durations, as Lerdahl and Jackendoff (1983) suggest for tonal music, can we accurately remember several durations at once.

Milton Babbitt modularizes duration with an entirely different approach, desiring "a system which will impose constraints upon the temporal elements of a composition with a comparable effect upon the nature and extent of the inter-event influence in the rhythmic domain to that of the pitch system in its domain" (Babbitt 1962). A duration series fails to do this: for example, combinatoriality fails because

---

[29] In other words, the set of all numbers n/m (where n, m are integers and m is nonzero) has more interesting properties than the set of all numbers m and $1/m$.

[30] His chamber work *Kontra-Punkte* hinted at this structure four years earlier, with a scale of tempos from 120 to 184 beats per minute. Maconie (1976) reports his intent in the earlier piece to construct analogies between pitch and tempo. Certainly the correspondence between change of tempo and transposition of pitch is immediately apparent to anyone who has used a multi-speed tape recorder or phonograph.

the two hexachords have unequal total durations. Duration is distance in time, as interval is distance in pitch; so Babbitt thought to interpret duration as interval instead of as pitch. Then pitch corresponds to the start of a temporal event, a time point. Modular structure is possible in a collection of time points, if the time points are embedded in a metrical unit such as a bar of 12 equally spaced time points (figure 3).

A time point is literally a point in time, imagining time as a line with a zero marked at the beginning of a composition and with marks of increasing value as the composition proceeds. If we modularize time into bars with twelve ticks each, we can identify a time point with its equivalence class modulo this bar—which sixteenth note of the 3/4 bar it lands on, irrespective of which bar of the piece that is. Given this definition of 12 time points, a *time point series* is defined as "a specific ordering of the 12 time points, extending across as many concatenated spans of modular length as necessary" (Mead 1994). The usual operations of retrogression, inversion, and transposition are defined for time point series just as they are for pitch class series.

Following Babbitt (1955) we can notate a series of pitch classes as a set of 12 pairs (order number, pitch class number), where both the order numbers and the pitch class numbers are each a permutation of the numbers 0 through 11. (This mathematical set is unordered; ordering is encoded in the explicit order numbers instead of implicitly by sequential presentation.) Using modulo-12 arithmetic, adding k to each pitch class number produces the operation of transposition by k semitones. Negating each pitch class number of a series produces its inversion; negating each order number produces its retrograde. We can now apply this rigorous definition of the elementary serial operations to series of things other than pitch classes: any collection of modes of attack, other pitch concepts, timbres, dynamics, registers, or whatever else we might wish to present sequentially, can also be presented serially. Even a collection of durations can be presented serially, within the time point system: on an instrument capable of sounding several notes at once, timepointing can specify attack points while this abstracted serial method independently specifies durations. (Note that series of length other than 12 are possible; no special properties of the number 12 are used in this abstract presentation, though it can be useful that 12 has many divisors. The number 12 is not prescribed by time, only by wanting analogous construction for pitch and duration.[31])

In this mathematical way of thinking about serialism, a series is fundamentally a sequence of values, in other words a function; from a discrete space to another discrete space to be sure, but a function all the same. Thinking of it as a fragment of a continuous curve, a melodic contour, this serial ordering becomes one particular way of composing with parameters. These curves are flipped, moved up and down, applied to pitches and time points and other parameters, in the hope that the composition will derive sufficient unity from using a small enough curve-*cum*-sequence (only twelve elements long). Even fewer than twelve elements can work: some of Schoenberg's predodecaphonic piano pieces have pitch material constructed of three- and four-note cells.

Babbitt (1962, 69) spells out the isomorphism between the systems of twelve pitch classes and twelve time point classes. Bar corresponds to octave; time point class to pitch class; temporal shift (metric reorientation of the series) to transposition; negating a duration (modulo the length of the bar) to inverting an interval (with respect to the octave).

A time point can mark the start or repetition of a pitch, chord, timbre, register, dynamic, and so on. Timepointing can be more than just the parallel presentation of pitch class series and time point series; the two structures of pitch and time points can be presented in the musical surface to emphasize either their relatedness or their independence as the composition demands. The composer Charles Wuorinen (1979) suggests several ways of using a time point series: it can be correlated with the same or another pitch class series; more or less than one new pitch class can occur at each time point; articulation or dynamic changes or instrumental changes instead of new notes can occur at time points; the modulus can vary (like Elliot Carter's metric modulations); and different time point series can be presented simultaneously. This last produces *resultant rhythms*, recognizably related to the component series if done with care.[32]

Westergaard (1965) critiques the use of timepointing in Babbitt's *Composition For 12 Instruments*. He finds the resultant rhythms of superposed rhythmic series to be muddled, unlike polyphonic presentation of pitch class series segmented by register or timbre. Combinatoriality does not work in detail anymore, as it disregards order within a hexachord. We can recognize a hexachord presented in any order, but only with difficulty a set of six attack points presented in arbitrary order. This is so because the ear accumulates pitches far more easily than it can accumulate slots for attack points and fill them in as attacks occur.

There is another difficulty with timepointing. Repetition of time points corresponds to octave doubling of pitches: both are a repetition of modules along the axis in which the modules extend. But repetition of time points *also* corresponds to repeating pitch classes, since repeating a pitch class in time is analogous to repeating a time point class in time. The isomorphism between serial operations on pitch

[31] Neither is 12 prescribed for pitch, if an alternative division of the octave is used.

[32] Steve Reich also uses the term resultant rhythm for superposed rhythms, in his case coming from phase shifts in compositions like his *Clapping Music*.

classes and timepointing breaks down subtly here, because the separate vertical and horizontal dimensions of pitch serialism (as drawn on a staff of music notation) collapse into a single horizontal dimension in timepointing. This is a loss of information, a noninvertible mapping. (Precisely isomorphic serial operations can be defined on any attribute of sound which is independent of the passing of time, such as loudness or register.)

In general, duration (or rhythm) stands alone, different from pitch and any aspect of timbre. Duration is unique among sonic attributes rather like philosophers consider existence as unique among all possible properties an entity may have. Music happens, fundamentally, in time. Even a silence takes time. All parameters of a sound, if they vary, can be said without loss of generality to vary *with respect to time*.[33] Time is unique here: we cannot describe all parameters as functions of, say, pitch or loudness. Duration asks to be composed with quite differently from other parameters like pitch, register, dynamics, or timbre.[34]

1.3.4 Dynamics

Dynamics, the loudness of sounds, is inherently imprecise. Electronics can accurately control the dynamic level of a sound, Boulez agrees; but this precision of dynamics may still be heard inaccurately by the nonlinear ear (Boulez 1963a, 60). In the realm of dynamics, "when composing, one should regard notation not so much as real than as, so to say, 'psychological'" (Boulez 1955, 54). It is trivial for electronics and difficult for performers to jump from *mf* to *ppppp* to *ffff* to *fff*. This is because of the instruments themselves: the loudness of most orchestral instruments is strongly cross-coupled with pitch. It is also because of the players' training: they rarely practise "arpeggios" of loudness on even a single pitch, never mind on varying pitches or with varying timbres. And even the coarse notation we have is traditionally relative to context: a deep *p*, a strong *mf*. But even if we restrict ourselves to electronic production of sound, we are still left with the psychoacoustic imprecision of perceived loudness. So "despite efforts to find a correspondence between intensity [*i.e.*, dynamics] and other dimensions of sound, it remains a necessary superstructure more than an actual organizing force" (Boulez 1955, 54). Stockhausen (1955, 61) notes that loudness is also cross-coupled with duration: louder sounds are imagined to be longer, quieter ones shorter.

---

[33] A parameter may not vary at all in a composition and still be considered important. For example, Stravinsky's *Symphony of Psalms* uses no violins, certainly not an accidental decision. But this merely jumps us up a level: the parameter of orchestration called "choice of string instruments" is still varying when we consider Western orchestral works as a whole. If something is called a parameter, it is thereby considered to be at least *potentially* varying on some scale.

[34] But pulsed rhythm is not the same as pure duration, pure passing of time. We understand an unchanging pulse to be a static parameter, and we are sensitive to variations in pulse as small as a few milliseconds. As long as a pulse is defined and its period remains in a limited range of about 0.1 to 2 seconds, its period can be varied with respect to time like any other parameter.

Boulez points out that Western tradition strongly correlates accelerando with crescendo and ritardando with diminuendo, but that Balinese music treats dynamics and tempo independently in these time-varying cases. This is ironic, since in physics we find that faster motion correlates with smaller range of motion. It may be explained as both speed and loudness being symptomatic of greater intensity in the basic musical alternation of tension and relaxation, and the sufficiency of a single measure of intensity for some musics. At any rate, decorrelating these parameters (diminuendo with accelerando, for example) then becomes powerfully expressive in a Western context.

Boulez sometimes notates dynamics indirectly, as in *Tombeau* from *Pli selon pli*. Instead of *ppp* or *mf* the score has a number (1 through 6) which is an index into a table of actual dynamics (*ppp*, *pp*, *p*, *mp*, *mf*, *f*). Several tables are defined; their entries are often monotonically increasing or decreasing. The performer is to choose one of these tables for a particular passage. But the composer could have chosen a table, too; then the score would have been notated conventionally. This is essentially serial construction, a series of dynamic markings instead of a series of durations or pitch classes; the table lookup corresponds to a reduced set of serial operations, inversion and a small number of transpositions.

A scan of the score of *Le marteau sans maître* reveals Boulez's two general ways of using dynamics. "Point-dynamics" are applied note by note, pointillistically (an overused but appropriate word). "Line-dynamics" are indicated by hairpin notations from one dynamic level to another. Either one or the other is in force throughout *Marteau*: the dynamics are explicit for almost every note.

Recall that Boulez's enumeration of blocks of durations ("symmetric irregular distributions," *etc.*) applies to dynamics as well. This can generate complex structures of dynamics. Indeed, Boulez correlates complexity of durations and pitches with complexity of dynamics.

Dynamics can have registers, by analogy with tessitura of pitch or tempo (Boulez 1963a, 64). These registers are realized as small-scale dynamic contours within an overall region of, say, *pp* or *fff*.[35] In practice these registers necessarily overlap, unlike the finely distinguished pitch continuum. (Overlap is also possible in dimensions other than dynamics.)

Effective use of dynamics is constrained by how the human ear masks quiet sounds with nearby loud ones. Variation of pitch, timbre (orchestration), and spatial position can ameliorate this effect.

---

[35] Mozart's *p* and *f* are often relative, in this spirit: a momentary *p* inflection within a prevailing *f*, or a true drop all the way to *p*. His notation is ambiguous by 20th-century standards because the scope of its dynamic markings is only implicit. In contemporary terms, Mozart's *p* and *f* sometimes indicate prevailing dynamic register and sometimes local variation.

1.3.5 Timbre

Boulez defines a musical instrument as “an ensemble constituted of timbres of limited evolution within a given tessitura” (Boulez 1963a, 64). This can be paraphrased as “a bounded collection of timbres of limited range.” Treatises on orchestration usually define the pitch range of an instrument, ways of playing it, and the sounds thus produced for a particular tessitura, mode of playing, and dynamic level. Instruments have historically been grouped into families, hence so too have timbres. But with the advent of *Klangfarbenmelodie* and *Klangfarbenreihe*, orchestration has matured from decorative to structural function (Boulez 1955, 55). The final chapter gives an elaborate example of such structural orchestration in my composition *COALS* (figures 37, 39, and 43).

In the early days of electronic synthesis, what timbres Boulez could create, he of course wanted to serialize. So to define a series of timbres we first need an organization of timbre that defines a distance, a step, from which we can construct a scale (Boulez 1955, 54). Schaeffer (1966, 372–374) confirms this method of constructing a scale from a continuum, adding that besides the step which gives variation to the continuum, one must explicitly specify what is invariant over the continuum. He demonstrates this in contexts of psychoacoustic experiments and of music, and repeatedly notes the paradox that “*ce qui varie, c’est ce qui est constant*”: having heard the variant part of the scale, eventually we hear it as constant (though on a higher level).

Boulez distinguishes three kinds of timbre: more or less constant, variable by steps, and continuously variable. Like dynamics, timbre is imprecise. With the electronics available in the 1960’s timbre was not susceptible to fine control—he states outright that timbre cannot be altered gradually (overlooking coarse controls like volume or filters, to be sure) although he looks forward to this possibility. So he has no discretized continuum of timbre to which he can apply serial procedures. Instead of a continuum, timbre consists of already complicated *objets trouvés*. So for Boulez the role of timbre is the “articulation” of pitch and dynamics, *i.e.*, giving these shape or coloring them in.

1.3.6 Spatial position

Spatial position of a sound source, unlike timbre, is eminently a discretizable continuum. Boulez imagines not only serialization of spatial position but also modules like octaves, an intermediate structure which divides the space into micro- and macro-structures (Boulez 1963a, 68). Of course room acoustics sully the geometric ideal.[36] (Compositions incorporating theatrical elements such as ritual processions use spatial position incidentally as part of a larger effect; we do not consider such here.)

Sound sources can be static or mobile, though Boulez warns that the theatrical effect of live performers moving about can distract from the sonic structures themselves.[37] Parenthetically, the visual presentation of even stationary performers strongly affects how a composition is received. Some say that visual presentation is the main reason pieces are written for tape with live performers. One composer writes:

> "...one of the most fundamental factors for engaging an audience into cognitive, intellectual processes has to do with seeing musicians in action. Let us take a caution to put an emphasis on seeing. The significance is not in the seeing itself. It is in the facilitation process of seeing the visual cues of performers' movement. These cues provide an intuitive access to the performers' kinesthetic energy control measured against and along the acoustic phenomena. Thus the seeing and the visual cues do not replace the listening. It is also not desirable if visual effects override listening experience." (Choi 1997, 25)

The importance of visible gestures of playing is also echoed by Favilla (1994): this is a major part of the design of his light-based LDR Controller. In fact, if new gestures are not visible enough to the audience, some *added* visualization of them may be justified. This additional visualization can also help in training the performer.

Implied spatial motion of a sound coming from first one place and then another can be heard as conjunct or disjunct—forming a single *auditory stream* or not, using the terminology of Bregman (1990). The percept of conjunct motion is enhanced by overlap of sound and by similar dynamics, timbre, pitch, or duration; motion may be conjunct by some of these measures and disjunct by others.

An early example of spatial control in the modern era is Boulez's *Poesie pour pouvoir* (1958) for tape and spirally disposed orchestra. Deliberate spatial structures were more popularized by Stockhausen's multi-orchestral works *Gruppen* and *Carré*; besides antiphonal use, fragile crossfades between orchestras was possible here. Boulez's more recent work *Repons* exemplifies the elaboration currently possible: six soloists and two sets of six speakers around the audience, together with an instrumental ensemble distributed among the audience. It is Iannis Xenakis, however, who most consistently and prolifically uses spatial positioning; we cite three pieces here. *Eonta* for piano and five brass players has the brass

[36] Human hearing is almost up to the task. We can resolve down to about five degrees of azimuth in the median plane (Sheridan and Ferrell 1981, 262). This yields a rough upper bound of $360 \div 5 = 72$ distinguishable values, apparently enough to justify dividing the spatial continuum into modules. But we cannot hear angles of azimuth nearly as intuitively as we can hear intervals of pitch; octave equivalence has no analog in spatial position. At any rate, headphones or anechoic chambers may be necessary for this level of refinement.

[37] Composers can use theatrical elements, of course; particularly Vinko Globokar and Maurizio Kagel have effectively combined contemporary instrumental composition with theatre. But unintended theatrics can harm a performance. Even watching the mechanical action of a player piano or the level meters on a tape recorder diverts attention from the sound itself.

moving around and even spinning while playing (a dramatic acoustic effect with trombones!), playing into the piano for extra resonance or away from the piano. Xenakis wished for players that could run while playing but met universal refusal on this point. Perhaps as a compromise he has worked with larger static ensembles, moving sound between players instead of along with an individual player. *Alax* situates three identical small orchestras around the audience and is notable for its use of spatial canons as well as "sound planes", each orchestra playing at its own constant dynamic to make sounds appear to come from the spaces between them. In figure 6, for example, one harp figure passes from orchestra III to II to I, while simultaneously a cello figure rotates in the opposite direction. Xenakis's larger orchestral piece *Terretektorh* arranges the players in eight sectors in a round hall, as individuals scattered throughout the audience. Circular motion of sounds through these sectors is prominent: reversals, spirals, and even "temporal spirals" where rotating sounds are gradually accelerated. Finally, with advances in software manipulation of sound, some recent electroacoustic works use spatial position as a primary organizational parameter: one section of Michael Clarke's *Prism* distributes the harmonics of eight trumpet tones among eight loudspeakers in various ways (Clarke 1999).

Figure 6. Bars 31–33 of Xenakis's *Alax* for three identical orchestras.

1.3.7 Intra-series structure

Twelve-tone series can be constructed to have internal symmetry. For example, some of Webern's series are a concatenation of a suitably chosen three-note figure with its retrograde, inverse, and inverse retrograde forms; Alban Berg's *Lyric Suite* opens with an all-interval series whose intervals' inversions are symmetrically disposed around its central tritone. Each internal symmetry defines a symmetry group (and an equivalence relation) on the 48 standard forms of the series. These symmetries can be exploited in the manner of pivot chords in tonal modulation, to move among different forms of the series.

Boulez sometimes limits the range of series, particularly series of duration and dynamics. He does this to avoid saturating the listener's attention towards the parameter of the series, which can happen when the range of this parameter is uniformly large. Wolpe goes farther, limiting the range of pitch class series to avoid the "grayness" of relentless use of the aggregate. Xenakis too was upset by gray sound masses resulting from dense complex polyphony; his book (Xenakis 1992) begins by stating that this was the impulse that drove him to use stochastics. We can generalize. In communications theory it is desirable to maximize the information carried by a channel. But this is not so in music, as a rule: *Structures Ia* shows what can happen when information density is high. Listeners appreciate some level of redundancy in music, if only to provide them with preliminary structures to parse and understand what they hear.[38]) In the presence of errors of transmission or reception, though, communications theory agrees that some redundancy is beneficial. In the case of music, performers may play inaccurately or lack some understanding of the music, and listeners may be distracted or inadequately prepared. It would seem a throwback to *Gebrauchsmusik* to compose with these constraints deliberately in mind, but it would seem foolish too to assume perfection. The composer does well to consider the results of human factors research, which field deals with the problem of how well a human can execute a task (performer) or observe a process (listener). Some of these results are discussed in section 3.3.

1.3.8 The spectrum of a series

If a series of twelve pitch classes is rendered in notes of only moderate length, from eighth notes to half notes, then the lengths of all these series will cluster around the length of twelve quarter notes, certainly within a factor of two of that. We can restate this by analogy with harmonic rhythm: the pitch class rhythm has a small range, less than two "octaves." I use the frequency-related term octave deliberately, as it suggests the idea of the spectrum of a parameter. Sound, a change of pressure level with respect to time, is usefully described by its spectrum of frequencies: the strength of each period of amplitude change, measured over some interval of time. This formalism applies to any parameter which varies with time, not just pressure level: harmony, pitch class, and dynamics are the most obvious. Wolpe essentially complains about narrow, almost "point," spectrums of pitch class. To go beyond a

---

[38] To be sure, ambiguity of parsing and specific cases of intentional unclarity provide much interest in music. As listeners grow more sophisticated, they need less redundancy and may even begin to consider it childish. For example, a music theory professor at the University of Illinois began a lecture: "I heard that Garth Brooks made 44 million dollars last year. That's about 15 million per chord."

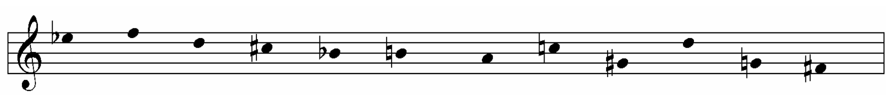
single shade of gray to many shades or even intense black and brilliant white, the spectrum needs to be broadened—to satisfy Wolpe, slower rates of change are needed. But this is true of any parameter: the composer must consider not only what values to choose, not only what ordering or other structure to impose on these values, but also the spectrum of change of values. A broad spectrum attracts attention to a parameter, a narrow one diverts attention elsewhere.[39]

1.3.9 Inter-series structure: extending a single series

Webern is known for exploiting a single interesting series in a composition. Boulez on the other hand excels at inter-series structure, that is, constructing a rich family of related series. Boulez (1963a, 40) alludes to what we may call his *table technique*, illustrated by (Koblyakov 1990, 3–6) with an example taken from the cycle of movements on *L'artisanat furieux* in *Le marteau sans maître*. (Later works by Boulez use variants of this.) First the rotations of the abstract series 2,4,2,1,3 are arranged as the rows of a matrix, choosing the sequence of rows so that the left column also reads 2,4,2,1,3:

```
2 4 2 1 3
4 2 1 3 2
2 1 3 2 4
1 3 2 4 2
3 2 4 2 1
```

Then each row is interpreted as a partition of 12, in particular as a partition of the pitch-class series used in *L'artisanat furieux*



so for example the fifth row yields the five pitch sets

[39] Looking ahead to the next chapter, the spectrum of an acoustic parameter powerfully indicates which controls of a synthetic instrument would work well with it. How often the parameter changes, how rapidly, with what accuracy—comparing these values with known values of controls (a joystick, walking to different places on stage, manipulating an array of keys or switches) can quickly simplify a messy problem of matching several controls to several parameters.

Let us notate this family of five pitch sets more generally as *a b c d e*. Now a table of all the ordered pairs of the five pitch sets is constructed: the elements of each row share the same first element, of each column the same second element.

aa	*ab*	*ac*	*ad*	*ae*
ba	*bb*	*bc*	*bd*	*be*
ca	*cb*	*cc*	*cd*	*ce*
da	*db*	*dc*	*dd*	*de*
ea	*eb*	*ec*	*ed*	*ee*

We write *ab* instead of (*a*, *b*) because this is actually a multiplication table. Each element of this table is produced by what is called *frequency multiplication*. For example, the product *ab* is built by constructing an *a* "chord" on each pitch of *b* and accumulating all the pitches thereby given. While Babbitt is content to deal with pitch classes in the abstract, Boulez inevitably presents them even purely didactically as having a particular vertical spacing, chosen so that no adjacent pitches form a semitone. This vertical spacing is required for the chord-stacking procedure. This multiplication resists formalization: in particular, the diagrams of pitch sets given in both (Boulez 1963a) and (Koblyakov 1990) erroneously assume commutativity. They actually represent the different table

aa	*ba*	*ca*	*da*	*ea*
ba	*bb*	*cb*	*db*	*eb*
ca	*cb*	*cc*	*dc*	*ec*
da	*db*	*dc*	*dd*	*ed*
ea	*eb*	*ec*	*ed*	*ee*

where the elements in the upper right triangle are simply copied from the lower left triangle. But even if this is a theoretical aberration, it does fit the notes of the music—after another modification to the table; the first row is arbitrarily (imaginatively?) replaced with *a b c d e*. At any rate, we end up with a 5×5 array of pitch sets, in fact five such arrays, one derived from each of the five partitions of the original pitch-class series. Koblyakov's term for such an array is a harmonic field; a collection of arrays is called a harmonic domain. Boulez then traces out paths through (again, possibly modified) harmonic fields to produce sequences of pitch sets used in the musical surface ((Koblyakov 1990, 10–22) provides examples).[40] This can be thought of as generating a vast collection of pitch-class series from an initial pitch-

[40] Such a path traced through an array exemplifies *traversal*, a common strategy for converting abstract constructions into linear sequences or contours directly renderable in the musical surface. The term is borrowed from computer science: an algorithm is said to traverse a data structure when it visits the elements of the data structure one at a time.

class series together with a set of partitions thereof: each row of the harmonic-field array is still an aggregate. Asymmetries are concentrated along the main diagonal when a path follows along it, because the individual elements `aa`, ..., `ee` have nothing in common. Series taken from off the main diagonal, in particular series taken from the rows and columns of the table, have intra-series symmetry by virtue of the presence of multiples of `a`, `xa` and `ax`; similarly for `b`, `c`, `d`, and `e`.

Frequency multiplication effectively creates a harmony more flexible and subtle than the predominantly intervallic harmony of dodecaphony. After all, a pitch class series is better thought of as a sequence of eleven intervals than as a sequence of twelve pitch classes. The classical operations on a pitch class series cannot change its sequence of intervals, even while they produce a variety of sequences of pitch classes; intervallic variation arises only in the combination and superposition of different forms of the series.

In the second cycle of *Le marteau sans maître*, namely *Bourreaux de solitude* and its three *Commentaires*, Boulez explicitly and rigorously correlates the parameters of pitch class, duration, and dynamic level/articulation by fixing certain dynamic levels, durations and articulation markings to the germinal pitch class series presented above. A similar elaboration produces a family of derived series which are actually used in the musical surface.[41]

In all this Boulez treats the series not as an ordered succession but as raw material to work with. The temporal succession of elements of the series is separated from the temporal succession of the musical surface.

It is instructive to compare this to similar constructions in Stockhausen's 1954 tape piece *Studie II*, analyzed in (Maconie 1976). Like Boulez, Stockhausen found it difficult at that time to directly manipulate individual sine tones to produce fused timbres. His solution here was to construct the piece from "tone-mixtures" of five sine tones; such a mixture was varied by means of filters instead of by (serially) specifying the amplitudes of its components.

In *Studie II* Stockhausen defines basic serial structure in terms of the number 5. He builds a scale of frequencies whose step is the ratio $\sqrt[25]{5}$, a little greater than a semitone ($\sqrt[12]{2}$). This scale extends approximately from 100 Hz to 5 kHz. Durations also have a scale based on $\sqrt[25]{5}$, covering about 50 steps. Tone-mixtures are built on frequencies 1, 2, 3, 4, or 5 steps apart on the frequency scale. The maximum intensity of tone-mixtures is governed by nine sets, each divided into five related sets.

---

[41] "Hierarchy" is Boulez's preferred term for such elaborations. The term is evident from the progression pitch – pitch set – harmonic field – harmonic domain – and even higher levels. This also explains why Boulez defines series itself as hierarchy.

The kernel of his construction is the series 24031.[42] He uses its transpositions—elementwise addition modulo 5—starting on 2, 4, 0, 3, 1 as the successive rows of a table:

```
2 4 0 3 1
4 1 2 0 3
0 2 3 1 4
3 0 1 4 2
1 3 4 2 0
```

(This construction is just like *L'artisanat furieux* except that it uses transposition, not rotation, of a series.) Stockhausen then builds a second table in the same way, starting with an elementwise inversion of 24031, replacing x with $4–x$ to get 20413. A row from one of these two tables indicates which frequencies are to be used in a tone-mixture. Rather like the "vertical" stacking of chords in Boulez's frequency multiplication, these rows are rendered as sets of five strictly increasing frequencies. Tone mixtures are presented in groups. In a given group, each tone-mixture has its own duration and maximum amplitude. A group then has one of three forms (or a retrograde thereof): (i) all tone mixtures begin together below the threshold of audibility, each one growing to its maximum amplitude and then stopping suddenly; (ii) the same, but muting part of the beginning of each tone-mixture; (iii) all tone-mixtures end together at their maximum amplitudes, therefore beginning at staggered times. Large-scale form is defined by a hierarchy of these groups.

In these two constructions we see the importance of symmetry and modular structure to both Boulez and Stockhausen. Babbitt, though, prefers the more rigorous technique of *partitioning* (colloquially, slicing out smaller things from a big thing). Two definitions are in order. First, by the *partition* of a number we mean a set of numbers which add up to it (all numbers here are positive integers). For example: 6 has partitions {1,1,1,1,1,1}, {1,1,1,1,2}, {1,1,1,2,2}, {3,3}, {6}, and twenty-nine others. We disregard order, so for example {1,2,3} and {2,1,3} are the same partition of 6. Second, the phrase "partitioning X by P" means to compositionally realize a particular division of the set X into subsets by assigning a different value for the parameter P to each subset. By listening to how P varies, we can hear the individuality of each subset and thus hear the partition. For example, we can partition pitches by register and instrumentation (or dynamics and articulation). In the diagrams of blocks of lynes which we have seen, this would mean assigning each lyne to a fixed register, instrument, or dynamic level. Wuorinen (1979) suggests partitioning a pitch class series into nonadjacent dyads of interval class 1,2,3,5,

[42] We renotate his 12345's as 01234's to clarify the mathematics.

or 6 (4 cannot partition an aggregate, as it turns out), and partitioning a single line into three or four voices while preserving register.

Mead (1994) describes the *all-partition array* which takes the idea of blocks of lynes (figure 5) even farther: its columns are all the possible partitions of a set of 12 elements into *n*-or-fewer parts for some fixed *n*. The series need to be constructed from all-combinatorial hexachords; those of first, second, and third order have respectively 3, 4, and 6 IH-related lyne pairs. We double these numbers to get the number of individual lynes, 6, 8, and 12. And the number of distinct partitions of the numbers 6, 8, and 12 yields the number of columns in an array: 34, 58, and 77 (depending, again, on the order of the all-combinatorial hexachord used to construct the series). Each partition then becomes a columnar aggregate, and each lyne pair is filled in with IH-combinatorial forms of the series. Lynes can be individuated by means of instrumentation, articulation, or register. (Chapter five illustrates instrumentation-based individuation of lynes in a serialization of trichords rather than pitch classes, in the composition *COALS*.)

These arrays are effective structural unifiers because they reduce arbitrariness: every note has several functions. Of course, arrays can be used to render time point series as well as pitch class series. Combinatoriality still holds here, so we can construct aggregates of time point series, for example in four lynes.

A *superarray* is a contrapuntal network of all-partition arrays or trichordal arrays.[43] For example, Babbitt's composition *Ars Combinatoria* starts with a four-block all-partition array. These four blocks, taken as a set, generate $2^4 = 16$ subsets. Discarding the empty subset, the 15 remaining collections of blocks are presented in four series.

1.3.10 Musical space

Boulez distinguishes continuous and discrete spaces with the terms *smooth* and *striated*. (A dictionary definition of striation applicable to this usage is: marked or streaked with narrow structural bands (Simpson and Weiner 1989).) Presented with a pure continuum, at a very low level we intuitively impose structure on it by imagining striations or features. Examples of this are the saccades of eye motion, seeing familiar images in clouds, and Western musicians' easy interpretation of highly microtonal scales as variations on conventional equal temperament. Formally,

---

[43] This is reminiscent of the elaborate heterophonies of polyphonies in (Boulez 1963a, 119–141), though Babbitt defines his methods more rigorously.

> "In practically all absolute judgment tasks, and especially in those involving only one dimension, some stimuli can be identified much more reliably than others because of their unique positions on the continuum. The effect is called anchoring. ...Anchors contribute information about the stimuli and result in more reliable judgments and hence higher transmission [of information]. The effect has been well analyzed by [(Garner 1962)]." (Sheridan and Ferrell 1981, 99)

By assuming that musical spaces have metrics, Boulez distinguishes the following kinds of striated spaces based on the structure of their striations. We commonly call striations modules. Modules of equal size induce *straight* spaces; modules of varying size induce *curved* spaces. Modules whose size varies according to some rule *focalize* a curved space around one or more *defining* modules, the *foci* (or anchors). A space may have submodular structure, a *temperament*: the obvious example is semitones within an octave-module. *Regular* spaces have constant temperament, even if the size of their modules varies (for example, sixteenth notes of constant duration within bars of varying length). Boulez imagines constructing instruments with controllable striation of pitch-space. I have prototyped this with the eviolin, rounding played pitch to the nearest tempered semitone to compensate for my own imperfect intonation. Rounding to other temperaments or scales is of course possible.

Smooth spaces, on the other hand, have no structure on which to build extensive analysis. Boulez distinguishes only two types: uniform and nonuniform probability distribution of values within the space. The latter he calls directed, presumably because nonuniform distribution at least leans towards structure (and in the limit of a discrete distribution becomes striated).

Boulez applies the terms smooth and striated to time, calling the two kinds of time *pulsed* and *amorphous*. He notes that counting, speed, tempo, and acceleration apply only to pulsed time; the only attribute of amorphous time is density of events (per unit time). But we note that density is measurable not in itself but only with respect to some constant duration. If we are given a fixed collection of a dozen events occurring in one minute and asked to draw the graph of event density with respect to time, we can legitimately draw any of the following curves: (i) a line zero everywhere except with points at infinity corresponding to the dozen events; (ii) a line of constant value, 0.2 events per second; (iii) lines somewhere in between, somewhat higher in the neighborhood of an event and somewhat lower in regions distant from all events. We can use a description of how density varies with respect to time, together with a random number generator, to define a sequence of events obeying this density distribution; but once we have this sequence of events, we cannot accurately recover the density distribution like we can recover a description of pitch versus time or amplitude versus time. This is another aspect of the special nature of time as musical parameter, which we saw while discussing timepointing.

Stockhausen tends to use continuous spaces while Babbitt and Boulez prefer discrete spaces. (Xenakis falls even more strongly into the continuous camp: his continuous gestures in pitch, dynamics, and spatial placement have given rise to more than one polemic from Boulez's pen.[44]) Unlike the often angular melodies of serialism and total serialism, a perusal of the score of the tape part of *Kontakte* quickly reveals a profusion of curves: continuous variation of pitch (22'–22'25"), of rotating spatial position among four loudspeakers (14'30"–15'30", 32'–33'40"), of amplitude (17'–21'), of bandwidth (23'45"–24'15"), of density (27'–27'45"). Individual excerpts are hardly worth illustrating, as these aspects of continuity are evident throughout the work. Stockhausen describes his 1956 tape piece *Gesang der Jünglinge* as "ordering sounds in a continuum of colors from sine tones to white noise" (Stockhausen 1955, 57). Maconie (1976) recounts how some of its evolving, nonstatic sounds were realized. Stockhausen constructed an instrument from a pulse-train generator, a filter, and an amplifier. This instrument was played by three people from a score of hand-drawn curves. One wonders how much the technology itself suggested continuity to him, as the apparatus he had in 1956 doubtfully could change values instantaneously but rather was controlled continuously with knobs.

Stockhausen observes the following phenomena occurring at various durations. From 1/13000 to 1/16 second, we hear timbre or noise color; from 1/6000 to 1/16 second we distinguish actual pitches, hence harmony and melody. From 1/30 to 1/16 second these percepts blur into duration ("meter") and rhythm. Durations are perceptible and distinguishable out to about 8 seconds; beyond that, we cannot clearly remember when the sound started. Finally, from 8 seconds to 15-to-60 minutes we perceive musical form. In summary, "each of the three large musical time-spheres—*frequency duration, rhythm duration,* and *form duration*—are of approximately equal size: each has a compass of about seven octaves" (Stockhausen 1962, 43). Not content with continuity of merely individual parameters, Stockhausen manages to build a single continuum from these three traditionally separate parameters by showing that our separation of them is perceptual rather than inherently physical. He demonstrates this physical continuum by carefully blurring and confusing its perceptual boundaries. For example, in *Kontakte* from 17'00" to 17'10" a pitch at F below middle C oscillates down and up a few times and finally descends below audible pitch, smoothly becoming a train of individual impulses (figure 7).

---

[44] "White and coloured noise, aleatory modulations, etc. … become unbearable just as quickly as their equivalents, clusters and glissandi—similarly amorphous and undifferentiated" (Boulez 1963a, 59). "Is space no more than a clockwise or anti-clockwise mannerism? The abuse of such *space-glissandi* seems to me to originate from an aesthetic just as summary as the immoderate use of clusters, glissandi, and other types of white noise…" (Boulez 1963a, 66).

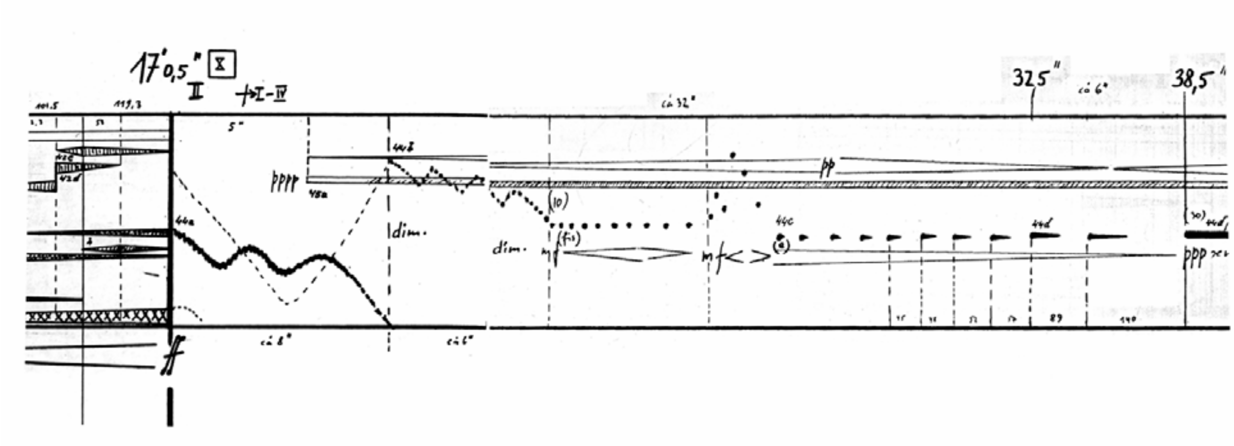


Figure 7. Excerpt from the score of Stockhausen's *Kontakte*. The descending curve starting at 17′ 0.5″ indicates a tone which falls below audible pitch and becomes separate pulses (drawn as dots) slowing down even more. Reproduced with permission from Stockhausen-Verlag.

This illustrates a more general principle: spaces can be juxtaposed (with respect to time or to some other parameter) or superposed simultaneously, to produce nonhomogeneous spaces.

Of course not everything in Stockhausen's music is continuous; he readily uses scales and discontinuous change of parameter value where appropriate. But the words he writes about continuity, contrasted with the polemics Boulez writes against it, are borne out in his music.

1.3.11 Critiques of Boulez's *Structures Ia*

Ligeti (1958) criticizes as simplistic the compositional process which underlies *Structures Ia*. He provides a recipe for creating such works: Choose your elements. Choose their arrangement. Choose an arrangements of these arrangements. Run the automaton defined by these choices. Finally and optionally, choose any unspecified parameters and choose to intuitively modify the output of the automaton.

Ligeti carefully notes that choice and lack of choice combine here; they do not oppose each other. Constructively rephrasing this rejection of dialectic between conscious decision and automaton, we can say that any work of art starts with some givens, some *objets trouvés*: pre-mixed paints, preconstructed musical instruments, established forms. This pattern is not restricted by using an automaton in composition. On the contrary, it is extended because the automaton is designed by the composer, not just plucked off the shelf.

(Many have criticized the use of randomness in composition since 1950. After half a century we would do well to discard the word random as so empty of meaning that all it can do is deceive. Merce Cunningham strongly defends the "random" combining of his choreography with Rauschenberg's staging and Cage's music, saying it led him to discoveries he could not have found otherwise. The essence of both Ligeti's automaton and Cunningham's randomness is the deliberate transfer of decision away from the composer, the hope of creating something unplanned.[45] Some consider this abdication; perhaps resignation is a better political metaphor, humble resignation in the face of being unable to proceed any farther under one's own power.)

From a series of 12 pitch classes Boulez constructs a matrix in the manner of the Boulez and Stockhausen examples presented above in the discussion of inter-series structure. He then replaces each pitch class with its order number in the prime form running across the top and down the left, in a sense normalizing the matrix with respect to the series used to construct it. He builds a second 12×12 matrix from the inversion (I0) of the prime form, and again replaces each pitch class with its order number in the prime form. These two matrices figure prominently in what follows. The asymmetry of construction (choosing order numbers not from I0 but from P0 in the second matrix) produces a dissimilarity in the two matrices. In particular the top row and left column of the first matrix read from 1 to 12 in order, while the second matrix reads as I0 in its top row and left column.

Dynamics are serialized with four intensity series built on the range *pppp* to *ffff*. These series are constructed from the main diagonals of the two matrices, a somewhat arbitrary choice but one that leads to symmetries in each series (a reduced series, Boulez would call it). Ligeti rightly points out that the precision of notation exceeds the precision of both performance and audition. Beyond this, ten modes of attack (sforzando, staccato, accented, *etc.*) are serialized (dubiously so, since they have no inherent ordering) using the opposite diagonals of the two matrices. But on pianos these modes of attack are difficult to distinguish in playing, never mind hearing, over an already superaudible precision of dynamics. On a percussion instrument, mode of attack is not independent of dynamic level. The one way in which mode of attack is audible is in the coarse distinction of staccato/legato: the score does not go so far as to notate a staccatissimo dotted half note, instead writing a short staccato note followed by a rest filling out the correct duration.

---

[45] One is reminded of Wotan trapped in his own laws, hoping that Siegfried will do what he cannot. "Der freier als ich, der Gott!" (*Die Walküre*, III). But just like the "random" free agent Siegfried does in fact set out to slay the dragon and thereby recaptures the ring of the Nibelung, the composer has some obligation after doffing Wotan the Wanderer's floppy hat to don Siegfried's helmet. The act of composition does not end with discovery but rather begins there.

"A compositional method exists only to write pieces. ...If the method has served long enough to allow the work it has produced to contradict it, it has more than fulfilled its function" (Wuorinen 1979). This seems to defend the contradiction between the construction and the perceptual result of *Structures Ia*. In hindsight this composition may have been a necessary failure, a *felix culpa*, to free serialism from its still tonal formal structures as practised by Schoenberg.

1.4 Formal models in the music of Xenakis

Though not explicitly named in a short essay by Xenakis (1955), *Structures Ia* certainly fits his critique of mid-twentieth-century serialism. He writes that serialism "destroys itself" in the contradiction between how its music is constructed (linear polyphony) and how it is heard (a mass of pitches and registers). All this technique is difficult to hear, so it is dispensable. Xenakis therefore rejected serialism and accepted—invented—other models of musical organization, from his background in mathematics. Vector spaces, sieves, and group theory are the main organizing principles running through his *oeuvre*. The same charge of inaudible technique has been levelled at Xenakis's compositions, as we shall see; inaudibility is obviously a subjective judgement! We can safely say, however, that choosing one of Xenakis's models is as fundamental a compositional decision as choosing dodecaphony. All these techniques—vector spaces, sieves, group theory, the various implementations of serialism—have one thing in common: they are mathematical structures imposed on a continuum.

1.4.1 Vector spaces

Xenakis (1992) defines a space of sound in terms of three axes, pitch, intensity (amplitude) and duration. By defining straightforward operations of vector addition and scalar multiplication, he calls the space of (pitch, amplitude, duration) triples a vector space. For example, the origin of the space may be set to (C3, 50 dB, 10 seconds), with unit vectors along each axis 1 semitone, 10 dB, and 1 second. This presents several difficulties when compared with conventional use of the term vector space. (i) This space is bounded: notes cannot be arbitrarily high or long or loud. (ii) It is asymmetrical: durations cannot be negative. (Negative amplitudes are avoided by measuring amplitude in decibels.) (iii) Scalar multiplication is perceived differently along each axis. Doubling the duration of a sound gives one the general feeling of a sound becoming twice as "much," but doubling its pitch (in number of semitones from a reference point, say C3) gives one the feeling of a sound becoming anything from slightly more (from C#3 to D3) to much more, even inaudibly more (from C7 to C11). Even worse, doubling the amplitude of a sound from 80 dB to 160 dB will give one the general feeling of pain and deafness.

(iv) The usable range and perceptual resolution is far from similar for all axes. (v) Our geometric intuition suggests that we might rotate vectors, apply rigid transformations to sets of vectors, and use other machinery from Euclidean space; but, as we have seen, the axes behave differently enough that a transformed vector is difficult to hear as related to the original. The structure of a vector space may find other applications in composition, but for the axes of pitch, intensity and duration, the combination of these axes is more accurately defined as a product of lines, without the additional structure of a vector space (scalar multiplication, vector addition, and certain properties of these operations).[46]

Kaper and Tipei (1999) formalize the concept of sound by removing time from Xenakis's space: "Sound is thus a vector-valued map from the time domain to sound space." This confirms our repeated observation that time must be treated specially among the parameters of music.

Alternatively, we can improve Xenakis's space by defining it in terms of log frequency, log amplitude, and *log duration*.[47] Take (C3, 50 dB, quarter note) as the origin point, and for axial unit vectors choose raising pitch one octave, increasing loudness by 10 dB, and doubling the length of a note. Then the basis vectors (1,0,0), (0,1,0), (0,0,1) have values (C4, 50 dB, quarter note), (C3, 60 dB, quarter note), (C3, 50 dB, half note). All three of these points feel "twice as much" as the point at the origin; the points (–1,0,0), (0,–1,0), (0,0,–1), "half as much"; the point (2,2,2) = (C5, 70 dB, whole note) "a lot more, maybe six times as much"; the point (–1,0,1) = (C2, 50 dB, half note) "about the same." Scalar multiplication still has the danger of quickly exceeding the limits of human hearing, so the origin should be chosen to lie near the middle of all these limits. Rigid transformations are now meaningful because "more" and "less" behave similarly along all axes. In particular, rotations can be used to interpolate smoothly between the different kinds of "more" which each axis offers. Translation of a set of vectors, adding a constant vector to all of them, generalizes three operations which can be done to a melody: making it louder or softer; transposing it to a different pitch; and augmenting or diminishing its rhythm.

1.4.2 Sieves

Xenakis uses his theory of *sieves*, a renotation of addition in finite groups, to organize collections of discrete points such as time points in a bar or microtonal scales. He uses the expression x_y or (x, y) to notate the residual class of y in $\mathbf{Z}_x$ (where $0 \le y < x$).[48] The additive group $\mathbf{Z}_x$ is implicitly a subgroup of some other predetermined group such as $\mathbf{Z}_{12}$ for conventional pitch classes or $\mathbf{Z}_{24}$ for the octave divided

---

[46] Others have used this three-dimensional description of sound without extra mathematical properties. Schaeffer (1966, 415) and many after him usefully represent individual sounds in a space given by hertz, decibels, and seconds.

[47] Log duration is suggested in (Boulez 1955, 51), although in a context of serialism rather than one of continuous spaces.

by equal-tempered quarter-tones. So in the context of $\mathbf{Z}_{12}$, 1_0 is the chromatic aggregate and 2_0 and 2_1 are the two whole-tone scales. (In the context of $\mathbf{Z}_{24}$, though, 2_0 and 2_1 are the two possible semitone scales, a quarter-tone apart.)

Using the set-theoretic operations of intersection, union, and complementation we can define any subset of the aggregate in terms of such sieves. Xenakis wisely presents expressions in the standard formats known as conjunctive normal form and disjunctive normal form.

This representation of a subset of a finite set is useful in certain contexts. Firstly and most rudimentarily, Xenakis (1990) concludes that sieves are essentially a language for describing the symmetries of a set of equally spaced points. Typically this set is embedded in a module which itself is regularly repeated: time points in a bar, or pitches in an interval (typically an octave but by no means necessarily).

Secondly, certain operations on subsets are easier to notate and manipulate by using sieve notation than by explicitly listing the elements of the subset. Given a pitch set described as a Boolean expression of x_y's, we can transpose it by n semitones by adding n to all the y's (the subscripts), or alter its intervallic structure by changing the y's individually. Such operations Xenakis calls *metabolae* or transformations of sieves (Xenakis 1992, 276). Analogous structures can be made at different scales by using the same expression but varying the embedded module ($\mathbf{Z}_{12}$ or $\mathbf{Z}_{24}$ or even larger ones derived from Byzantine music, or varying bar lengths); analogous structures can be made in pitch, in time, and in other parameters—this is Babbitt's longstanding desideratum. Sieve notation is unwieldy for small sets, but becomes progressively more powerful the larger the embedded module gets. Tipei (1999) gives an example of 120 bars of 4/4 time where each beat can be divided into quarters, fifths, sixths or sevenths. The least common multiple of {4,5,6,7} is 420; $420 \times 4 \times 120 = 201600$ attack points. Using formulas instead of lists is obviously desirable here.

Finally, within the domain of duration sieves are more general and flexible than timepointing; in parameters beyond duration, sieves are a more rigorous formalism than Boulez's generalized serialism. Sieves may be more appropriate than the array constructions of Babbitt when significant symmetrical structure is present.

Weighted sieves are to sieves what fuzzy sets are to sets. Instead of defining fixed subsets of the aggregate set, a weighted sieve associates a probability between 0 and 1 to each member of the aggregate set. Set-theoretic operations on weighted sieves are then defined as they are in fuzzy set theory. The music composition program MP1 uses weighted sieves to define pitch sets with precise probabilities for

[48] The standard mathematical notation for x_y is $[y]_x$.

the occurrence of "nondiatonic" pitch classes outside the set, and also to define distributions of attack points in a bar which approximate conventional ideas of weak and strong beats (Tipei 1981; Tipei 1999). Weighted sieves can be thought of as a formal arithmetic applied to discrete probability distribution functions, an extension of stochastics analogous to Boulez's extension of dodecaphony by frequency multiplication.

1.4.3 Group theory

The compositional technique of Xenakis developed from sound *masses* in stochastic works like *Pithoprakta* (1955), proceeded through sound *states* in the form of Markov processes, and finally arrived at a more developed *control* of sound states by exploiting the area of mathematics known as group theory (Vriend 1981, 42). The two compositions best exemplifying his use of group theory are *Nómos* α for solo cello and *Nómos* γ for orchestra distributed among the audience. We now consider the first of these in some detail.

DeLio (1980) explains the structure of *Nómos* α well beyond what Xenakis achieves in his own discussion in (Xenakis 1992, 219 *et seq.*). The whole piece is divided into two levels presented as 24 interleaving sections: three sections of level 1 followed by one of level 2, this pattern repeated six times in all. (The transitions between levels 1 and 2 are articulated by a change of pitch set; these twelve pitch sets are related to each other and defined by means of sieves, of course.) Xenakis then takes the group of symmetries of a cube, called the hexahedral group, and two normal subgroups thereof.[49] He constructs paths through the group by what DeLio calls Fibonacci motion: the path begins with two arbitrary elements x_0 and x_1, and is continued according to the rule $x_i = x_{i-2}x_{i-1}$.[50] (Figure 8 shows this motion for a group smaller than the hexahedral group.) Graphically, each multiplication by x_{i-1} to get x_i corresponds to rotating the cube onto itself, about some axis going through the midpoints of opposite faces, through the midpoints of opposite edges, or through opposite vertices. Since the hexahedral group is finite, these paths eventually terminate in cycles.

Two of these paths through the group are presented simultaneously, used to render different parameters. The first path is indirectly rendered in terms of octuplets of *sound complexes*; the sound complexes are eight gestures of playing such as a cloud of points ordered and ascending, a stochastic field of glissandi, or sustained sounds at constant pitch. For each interval of time, its associated group element x

[49] Delio's article introduces enough group theory that the novice can grasp his later arguments. (The hexahedral group has order 24 and is in fact isomorphic to S_4.)

[50] This name alludes to the familiar recursive definition of Fibonacci numbers, $x_i = x_{i-2} + x_{i-1}$.

is written as a permutation of (12345678); this permutation can be thought of as the new positions of the labeled eight corners of the cube. The resulting order of the numbers 1 through 8 is interpreted as a sequence, and is sent through a mapping to produce a sequence of eight sound complexes (gestures). To increase variety, Xenakis uses not just one but three mappings from number to sound complex. Each of these mappings more or less separates the staccato from the legato gestures, so a bipartite structure can be heard in the presentation of each group element. Another formal symmetry governs which of the three maps is used as the piece progresses. The second path is used to define the values of three parameters of each sound complex: density (attacks per second), loudness (*mf* to *fff* for level 1, extreme dynamics for level 2), and duration (seconds). Each of these three parameters can take on one of four discrete values. It is worth noting that Xenakis uses these three primary sonic parameters, density, loudness, and duration, together instead of applying independent structures to each one.

·	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

(a)

$$4\cdot2=1;\ \ 2\cdot1=3;\ \ 1\cdot3=4;\ \ 3\cdot4=2;\ \ 4\cdot2=\ldots$$

$$\Downarrow$$

$$(4, 2, 1, 3;\quad 4, 2, 1, 3, \ldots)$$

(b)

$$0\cdot1=1;\ 1\cdot1=2;\ 1\cdot2=3;\ \ 2\cdot3=0;\ \ 3\cdot0=3;\ \ 0\cdot3=3;\ \ 3\cdot3=1;\ \ 3\cdot1=4;\ \ 1\cdot4=0;\ \ 4\cdot0=4;\ \ 0\cdot4=4;$$
$$4\cdot4=3;\ \ 4\cdot3=2;\ \ 3\cdot2=0;\ \ 2\cdot0=2;\ \ 0\cdot2=2;\ \ 2\cdot2=4;\ \ 2\cdot4=1;\ \ 4\cdot1=0;\ \ 1\cdot0=1;\ \ 0\cdot1=1;\ \ 1\cdot1=\ldots$$

$$\Downarrow$$

$$(0, 1, 1, 2, 3, 0, 3, 3, 1, 4, 0, 4, 4, 3, 2, 0, 2, 2, 4, 1;\quad 0, 1, 1, 2, \ldots)$$

(c)

Figure 8. "Fibonacci" path through the group $\mathbf{Z}_5$. (a) The operation table of the group. For example, $2\cdot3 = 0$. (b) The path starting with 4 and 2. (c) The path starting with 0 and 1.

Vriend (1981) is even more exhaustive than (Delio 1980), having had the advantage of regular correspondence with the composer while conducting his research. He questions how appropriate it is to map the eight corners of a cube to a sequence of eight sound complexes, since the correspondence is hardly audible. Certainly we can say that there is no question of possible isomorphism, as the mapping is between such different domains: the vertices of a cube have no inherent order of traversal like a sequence does, and the elements of a sequence do not have the rich spatial relationships of adjacency and relative distance which the vertices of a cube have. Vriend (1981, 31) draws a generalization from this:

> "[W]e can choose an extra-musical mechanism because it guarantees and controls a special type or degree of *variation*. If we consider music as a chain of sound-variations in time in all its dimensions (pitch, intensity, density, degree of order, timbre *etc*.), a continuous flow of sound fluctuations as it were, we need a regulating mechanism to control and to steer it. Group transformations are a mechanism of variation from this point of view, and the composer should take care the construction of his states is best suited to the mechanism he chooses."

In other words, group structure is a constraint, a set of symmetries, imposed on the combinatorial explosion of possible values of parameters. In the case at hand, (Vriend 1981, 74) reinforces Babbitt's warning:

> "The 'logic' supposedly present in the chains of group transformations is not simply transplantable to questions of logic in the domain of a listening strategy: for example, if we would like a listener to be able to follow the string of transformations [the Fibonacci path] and, when completed, to realize that a 'loop' is closed at the end... that would not be a fair, or adequate problem: it is not sufficiently stated in musical terms, it is not sufficiently a musical problem."

We should mention in passing that what DeLio calls two layers Vriend calls two paths, since each is constructed from a particular traversal of group elements. We hold to DeLio's term for clarity, though. Vriend also remarks that the sections of layer 2 are visibly less developed than those of layer 1; he even considers them "intermezzi" between the larger sections found in layer 1. Going farther, (Vriend 1981, 41) presents another method of choosing two paths which are more closely coupled in the structure of the hexahedral group, demonstrating how the composition could have had a tighter unity. Vriend has in fact composed works with this more refined group traversal.

Recall that the second path is used to define the density, loudness, and duration of each sound complex. Vriend (1981, 34–36) reports that originally twelve, not four, values of density, loudness, and duration were defined.[51] A weak modular structure was imposed on each parameter: the twelve values were divided into three modules of four elements each. At any particular point in the piece, we find in force one module for each of density, loudness, and duration; this gives 4×4×4 = 64 possible values for the (density, loudness, duration) value applied to a sound complex. Xenakis takes care to not repeat such values when applying them to successive to each of the eight sound complexes in a sequence, to increase variety in the musical surface. (Vriend (1981, 39) notes that the variety generated is not nearly as great as it might be: only a small part of the 4×4×4 cube of values is explored.)

The density referred to above specifies the number of events per second within a single sound complex. So at the smallest scale the listener should be able to individuate events, as well as individuating the sound complex containing these events. Xenakis ensures this by varying parameters continuously within an event but introducing discontinuities in at least one parameter at the boundaries between events (Vriend 1981, 52).

A constraint between the two parameters of tessitura and mode of articulation (*battuto col legno*, *pizz. glissando*, *etc*.) is defined by what (Xenakis 1992, 228) calls kinematic diagrams. Here neither tessitura nor articulation depend one on the other. Rather, both are functions of a third abstract parameter. Lest this usage be confusing, the word parameter here alludes to the method of defining curves which is called parametric equations: $x = f_x(t)$, $y = f_y(t)$ instead of $y = f(x)$. A grid of 4 tessitura-ranges by n modes of articulation is traversed one cell at a time; this traversal is then the parameter t moving through its range and thereby specifying x and y. Traversals of the grid are chosen so that successive cells of the traversal are either maximally similar or maximally different (Vriend 1981, 68).

Xenakis defines pitches with sieves based on the prime numbers less than 18 and not a factor of 18 (so excluding 2 and 3): 1[sic], 5, 7, 11, 13, and 17. These sieves are presented as large grids. Vriend could not rigorously reconstruct these grids based on the score:

> "And to top it all, Xenakis admits that he made 'adjustments' as well as errors. Adjustments, because he sometimes 'judges' a grid to be *not interesting enough.* Criteria for 'interesting' are purely intuitive, he states, and so we are left once more with confusion and start wondering whether the construction of grids cannot be made more simple and more close to the composer's requirements. ...If one goes as far as this, one would be far better off with a stochastic approach, which can easily be adjusted without interfering too much with the essentials of a distribution." (Vriend 1981, 64)

---

[51] Xenakis could not reconstruct the exact sets of twelve values fourteen years later, so Vriend estimated their values from statistical analysis of the score. Vriend found that the values of both density and duration were logarithmically scaled in practice, unlike the linear scaling for vector spaces preached in (Xenakis 1992).

More generally, it seems safer for composers to use the simplest formal system they can get away with for choosing parameter values. If a big mathematical hammer is used, it should pound big nails.

Finally, we should note this common technique of *traversing* a structure in order to render it in unidimensional time. Here we have seen traversal through a group table defining large-scale form and traversal of grids of tessitura versus articulation defining details of the musical surface. Recall too Vinko Globokar's (continuous) traversal of a dodecahedron in *Les Emigres*. Interesting formal structures by their nature tend to be multidimensional; such a structure can be mapped to an individual aspect of a composition, a single varying value, by following a path through it. The choice of path powerfully influences the final result, though; some fraction of the effort spent defining a formal structure and its mapping to sound is therefore appropriately reserved for exploring different paths through it.

1.5 Indeterministic control of parameters

Aleatory music is stochastic, deliberately unspecified or unchosen or undetermined at some level of detail. The details are filled in either by the composer (perhaps delegated to a computer) or by the performer (perhaps delegated to the "presets" of the instrument designer). Aleatory composition inevitably brings parameters into conscious thought, because a distinction must be made between what is specified and what is unspecified in the score, and again because some constraints must still be placed on the unspecified aspects (such as their range or rate of change).

The simplest indeterministic control of a parameter is assigning it a probability distribution and letting it go at that.[52] This is the classical stochastics of early Xenakis. Temporal structure can be added to this by letting the value of the parameter follow a random walk or follow the sequence of distributions given by the transition table of a Markov process. These last two are examples of random processes with memory, ways of (discretely or continuously) varying a parameter which deliberately acknowledge that the parameter is heard as varying in time, not viewed statically as a graph printed on paper.

Some indeterministic descriptions are better implemented with manually executed algorithms, others with a computer, others by the performer; each has particular strengths and weaknesses. Historically there has been a contrast across the Atlantic in where indeterminacy is applied to music, whether the performer or the composer rolls the dice. European composers have tended to deterministic composition with indeterministic performance—Krzysztof Penderecki's sound masses, Stockhausen's *Klavierstück XI*, Boulez's later piano works. Americans however prefer the reverse, indeterministic composition with

[52] (Lorrain 1980) is an excellent statistics reference applied to music.

deterministic performance—John Cage's *Music of Changes* and Dick Higgins's unsurpassable *Thousand Symphonies* of machine-gunned blank orchestra manuscript (Cope 1971, 91). On both shores we also find a small amount of indeterministic composition *with* indeterministic performance—conceptual music, merely a few often mystical and cryptically poetic instructions to the performers. Little has been written about these compositions, notably some later works of Stockhausen and Cage; little *can* be written. If talking about music is like dancing about architecture, then talking about Stockhausen's *Aus den Sieben Tagen* is like dancing about building permits.

1.5.1 Free stochastic music

We have already mentioned Xenakis's string orchestra work *Pithoprakta*, where the speed of glissandi corresponds to the velocity of molecules in a gas, obeying a Gaussian distribution. Defining the value of a parameter with nothing more than a probability distribution function is the most elementary use of indeterministic control.

Xenakis's composition *Achorripsis* uses a more elaborate structure. The piece is graphed out as a matrix of cells, time along the horizontal axis and a variety of timbres along the vertical axis. The number of events in each cell, 0 to 4, is given by a Poisson distribution. The temporal location of events in a cell (each cell is about 15 seconds long) is distributed uniformly.

1.5.2 Markov processes

To extend his stochastic music beyond static probability distributions Xenakis began to use Markov processes, a formalism which can be described as a probability distribution with temporal behavior. Indeed, for discretely valued parameters an order-zero Markov process is the same as a fixed probability distribution.

Xenakis (1992, 79–109) describes his application of Markov processes directly to a sound space. He divides a space marked by axes of frequency, amplitude and time into a very fine lattice. Each cell of the lattice is small enough to contain a "grain" of sound; clouds of these grains produce more elaborate sounds. A single time slice of this lattice is called a *screen* of frequency and amplitude. He applies set-theoretic operations to simple screens (bounded areas, lines, sets of points) to produce other screens. Markov processes of these screens then produce the final acoustic surface. Xenakis (1992, 94) notes that as a low-entropy Markov process may tend towards an equilibrium state, musical variety can be increased by artificially perturbing the process; Boulez would call this local indiscipline.

Xenakis used this formalism in 1959 in two works *Analogique A* and *B* for strings and sine waves respectively. For each of the parameters frequency, amplitude, and density (notes per second), he defined two complementary Markov tables F_1, F_2, A_1, A_2, D_1, D_2. At each point a higher-level Markov process determined which F_i, A_i, and D_i to use.

1.5.3 Quasi-deterministic control

Values for a musical parameter may come from numerical data in nonmusical domains: geography, astronomy, simulated fractal processes, stock market data, social security numbers, and activity logs of computer users give an idea of the breadth of possibility here. Even sound itself can be remapped to sound, in fairly literal ways as with *musique concrète* manipulations and Messiaen's famous bird songs or more subtly as with the translation of timbre into complex rhythms in Stockhausen's *Gruppen* for three orchestras. Maconie (1976) mentions three of these translations in *Gruppen*: a sharp pizzicato becomes a burst of short high notes, the weak fundamental rendered as oscillating dynamic level; a sustained note after some attack transient sounds becomes a steady oscillation; a woodwind note with slow, smooth onset becomes an ostinato rhythm starting simply and building in complexity.

The sheer variety of domains which can be mapped to musical parameters suggests that many mappings are uninteresting. Babbitt's question returns: what makes a domain "appropriate for interpretation as a metrical or ordinal pitch, or durational, or dynamic, or timbral, or density determinant?" (Babbitt 1972a, 7) Certainly the first answer to this question is found in the spectrum of the process in time (mapped from the raw numerical data), as we discussed in considering the inter-series structures of Boulez. If the process's rhythms and rates of change of value fit well with the parameter being rendered, then we can call that connection between process and parameter "appropriate" in Babbitt's sense.

The source of the data, strictly speaking, has little relevance to the answer. Even with hints from a program note or a title (Larry Austin's *Canadian Coastlines*, Charles Dodge's *Earth's Magnetic Field*) the audience will have difficulty identifying the original data.[53] After all, this is composition and not merely the sonification of data. The source of the data may inspire other choices which the composer makes, but this is beyond the scope of the question: is this data appropriate for this musical parameter?

Looking backwards at the question, the composer might begin by wanting to construct a musical analogy of some domain represented by a set of data. Given the data, then, what musical parameters can we map it to? (We assume that the mapping should preserve structure. Interesting results can happen

[53] Stockhausen too has indulged in geography: (Stockhausen 1989, 45) claims that many of the rhythms in *Gruppen* derive directly from the outlines of mountains in the Swiss Alps.

when structure is whimsically or dramatically distorted, but such results are hardly predictable never mind formalizable.) The first aspect of the data to consider is its resolution. If interesting structures lie in fine-grained detail of the data, but the data nevertheless has a broad range, then this structure will be best preserved by mapping it to a finely discriminated parameter like pitch. If the structure of the data is coarser, pitch may be better reserved for things which need such fine resolution. If the data is essentially continuous in nature, even if sampled discretely, its structure is more clearly heard in mappings to continuous parameters like loudness or timbre, not discrete ones like equal-tempered pitch. Clustering of the data should also correspond to clustering possibilities in the musical domain: if the total range of values is covered fairly uniformly, a parameter like duration or time point works well; if the data is visibly clumped, fuzzy discrete rendering (as defined in the introduction to techniques of parametric organization) into "terraced" dynamic levels suffices to preserve the essential structure of the data.

1.5.4 Sets of rules as a score

Both the performer and the composer contribute something to a performance. In traditional Western repertoire the performer contributes varying timbre and intonation, deviations from precise meter, and idiomatic instrumental technique (fingering, choice of hand position, *etc*.). These things are thus not notated exhaustively by the composer. Total serialism tried to claim all this for the composer, notating everything, leaving as little as possible to the "whim" of the interpreter. Behrman (1976) suggests that the best a performer of such works can do is to be a non-interpreter. But this works only to a point: the more total the total serialism, the more exact the notation, the less likely the performance will be accurate. The limited precision of human neuromusculature (to say nothing of attention or listening) made the pure goal unattainable. Behrman quotes Cornelius Cardew taking the opposite stance: musical notation "should be directed ... towards the people who read it, rather than towards the sounds they will make."

Some composers explored this opposite direction, embracing the weaknesses and strengths of the performer by designing notations following Cardew's dictum. A simple example is the "race-course" form of Morton Feldman's *Durations I* and Frederic Rzewsky's *Les Moutons de Panurge*: players start together, proceed independently, and stop when they individually reach the finish line. A similar result could be achieved through conventional notation, but would require much more rehearsal and would threaten to misdirect the attention of performers (and listeners and theorists) to irrelevancies, particular simultaneities and sequences. John Cage once defended the use of chance (by composer or by performer) by saying if he wanted to show you his collection of marbles he should just empty the bag onto the table, not carefully arrange the marbles in a pattern. This is the thinking behind notations like Feldman's specification of pitch as only high, middle, or low: pay attention elsewhere.

Christian Wolff has composed a number of pieces where the score is far removed from the conventional sequence of parameter values (a tracking task, in human factors terminology). Instead of executing certain gestures in a defined order and rhythm, the performers make gestures in undetermined order but subject to strict rules which vary in their details from gesture to gesture.[54]

Wolff's notations can be divided into three families, those for a single musical event (a note), those describing the passing of time in a single event, and those defining temporal relationships between events.

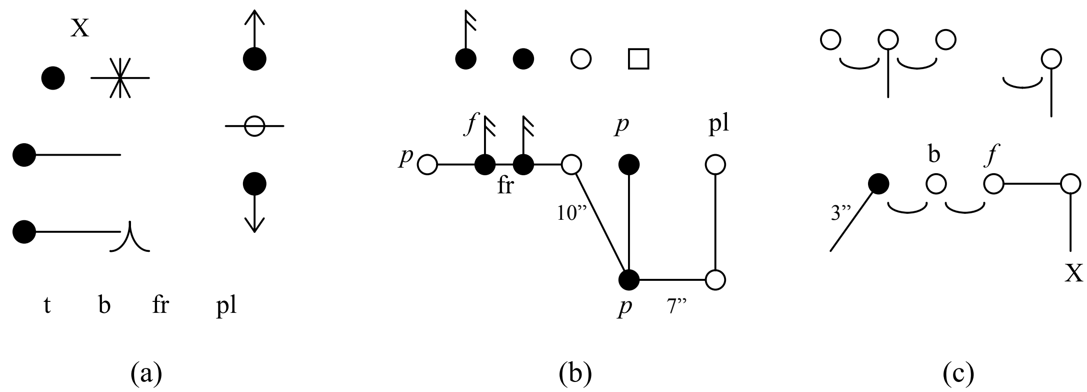


Figure 9. Notations used by Christian Wolff. (a) Single events; instructions idiomatic to an instrument. (b) The passing of time in events of a single player. (c) Temporal relationships between events performed by different players.

The simplest notations do not relate to the passing of time. As with other composers, a large brace containing several fragments indicates that the performer should arbitrarily choose one of the fragments. This is the fundamental level of unpredictability in the musical form produced by this notation: choosing one of several possibilities. In a fragment, a particular number of notes (possibly zero) may be specified to be played from a collection of notes.

Figure 9a illustrates particular notations used for idiomatic playing techniques (sul tasto, playing inside the piano, *etc*.). An asterisk is a nonspecific indication for a noise produced by means other than

---

[54] An elaborate, even baroque, instance of this technique is Stockhausen's 1963 composition *Plus-Minus*.

conventional playing technique. Where idiomatic technique does not apply because instrumentation is unspecified, as in *For 1, 2 or 3 People*, short abbreviations indicate sounds made by tapping, breathing, friction, plucking; other notations indicate extension, cutting off, raising or lowering the sound "in some aspect." A number in square brackets means that that many notes should be played together (on a keyboard, typically). If *a* is the name of a set of pitches, then "*a* ½" asks for playing a note a semitone up or down from a note in *a*. The notation explores the full range from fixed through constrained to free choice of both pitches and dynamics. Some constraints are temporal: the symbols *x*, *y*, and *z* found in some of the fragments in Wolff's *Duo for Violinist and Pianist* indicate a relationship between pitches. An *x* is free, a *y* must differ from the preceding *x* (whenever it was), and a *z* must differ from the preceding *x* and *y*.

Notations related to the passing of time of a player's own sounds require more attention from the player (figure 9b). Sixteenth-note, black, white, and square-white note heads indicate notes which are respectively very short, less than one second, unconstrained in length, and very long. Lines connecting note heads indicate temporal correspondence of the player's own sounds: a vertical line indicates starting and stopping together, a diagonal line that the second should start when the first stops. A duration in seconds of an entire fragment, or part of a fragment, may be specified.

The richest musical interest comes from lines and slurs indicating the correspondence of the start or end of a sound with the start or end of a sound from *another* player: of the next sound heard, or of the next sound heard of a particular kind (constrained to have a particular dynamic level or other attribute). Various notations indicate that the player's sound should start (or end) before, together with, or some (particular or arbitrary) duration after the other player's event has happened. Other notations use slurs to indicate compound constraints of this type, such as "start anytime, hold till another sound starts, continue holding anytime after that sound has stopped" or "play after a previous sound has begun, hold till it stops" (figure 9c). A much milder form of this is found throughout the musical surface of *COALS*, as extended runs of grace notes which cannot literally be played in the time allocated to them on the page. The performer of such a passage has to explicitly decide where to steal the time from, which synchronizations to make and which to abandon with the lines of the other performers. The rule in *COALS*—play these notes quickly yet clearly, and get back into tempo afterwards—sits in the middle ground between Wolff's rich landscape and the even smaller-scale rules for playing individual grace notes in common-practice music.

We also see that music structured by such rules cannot be synchronized by a conventional conductor like *COALS* is. A conductor might act as rule-interpreter, separating the tasks of listening and playing;

this would make playing easier, but it would also reduce how "tight" the music is by adding an extra layer of human reaction time. Behrman (1976) vividly compares the result of such rule-based notation to returning a fast serve in table tennis, and the notation itself to "rules governing the conduct of games." Compared to a conventional score this notation is complicated, because of the variety of things the player must think about *while playing*. The resulting musical surface (or game play), simple yet intense, contains grace notes between players, long notes cut off after another one starts, and an overall uncertain feeling from the players. We can feel them on edge, trying to rigorously wait for a particular sound from another performer, freely decide what choices they will make in the next fragment, and precisely remember what the last two x and y pitches were.[55] Behrman finally justifies this notation because of the great care it causes the players to take, listening to each other lest they lose their place.[56] "Music must remain a creative activity for players as well as an arrangement of symbols by the composer."

In these compositions Wolff is thinking in terms of parameters: duration, density (notes per second), pitch, pitch class, register. He does not control timbre so strongly, preferring to leave the details of its discrete colors to the performers.

Stockhausen's composition *Spiral* for one soloist also divides the score into score *per se* and a detailed legend of notations, in this case a single page of score and ten pages of explanation. One noteworthy device is his use of plus, minus and equals signs to indicate more, less, or the same of some parameters in a sonic event as compared to the previous event. Exactly which parameters are to be increased or decreased (from the fixed list of duration, register, dynamics, and division into shorter notes) is left to the performer.

1.5.5 Intentional and accidental errors

There is a third way to control parameters. Besides being controlled deterministically or indeterministically, they may be (at moments, at least) uncontrolled. This may be the only truly random way of composing, beyond even Cage's aesthetics: the composer, rather than merely wishing for sounds to be free of his intentions, does not even intend to not intend something.

---

[55] It seems impossible for a human conductor to follow several of these rules at once, for a group of performers. A computer with a microphone and video display for each performer and a reliable gesture recognizer might be capable, though it would take some of the fun out of the performance.

[56] A similar intensity and on-edge-ness is extracted from performers by the works of Brian Ferneyhough, but from the opposite direction: precise but almost impossibly intricate conventional music notation. The common element between Wolff and Ferneyhough lies in their challenging the performers.

Ligeti discovered several minor disagreements between the formalism and the score of Boulez's *Structures Ia*, perhaps intentional, perhaps accidental. "In strictly regulated composition, 'gratuitous actions' are conceivable, even desirable; by introducing the unpredicted into the predetermined, extraordinary artistic effects could be produced. But an inaudible act of arbitrariness will shock nobody—except maybe the analyst" (Ligeti 1958, 47). Ligeti even proposes to submit error itself to serialism, in a scale ranging from error-free through nearly inaudible to major deviations.

Similarly, a theorist in analyzing a composition constructed on an elaborate system by Xenakis disclaims:

> "We shall encounter several things in the score which do not precisely fit the data. Xenakis explains this [in personal communication with this writer] by saying a) 'in the heat of the action I made slips of the pen which I discovered only too late, after publication (of the score as well as his book – JV); b) I sometimes change details because they appear to me more interesting for the ear and c) I make theoretic errors which entail errors in the details." (Vriend 1981, 44)

This idea of embracing errors persists in classical music despite the note-perfect aesthetic suggested by commercial recordings; here are two examples.[57] In 1998 I asked the composer Brian Ferneyhough why, after carefully designing a system for a piece, he worked in ink and allowed such inevitable deviations from the system, errors, to creep into his work. He answered that he felt there would be some reason why this error and not that error had occurred, and wanted to honor what actually made its way onto the page. And performers too can push themselves past perfection: piano professor William Heiles maintains that a Beethoven sonata is not played right unless the pianist misses a few notes. This is not a problem with Brian Ferneyhough's music, though: Feller (1994) reasons that Ferneyhough's music is too difficult to play perfectly, so what performers contribute to the performance are *failures*, their own versus those of someone else.

1.6 Conclusions

1.6.1 Individual parameters

If an extramusical model is chosen so that a structural analogy exists between the abstract object and the musical parameters subjected to it, then the model will be more clearly heard and thereby justified as source material. The more directly acoustic the musical parameters are, the stronger this connection is. If

[57] Other musical traditions have freer attitudes to errors. Ornette Coleman is reputed to have said that jazz is the art of riffing on your mistakes, while Schoenberg is often quoted as pointing to an eraser and saying "this is the most important end of a pencil."

the music is performed with a synthetic instrument such as the eviolin, such strong connections can be more easily made with the help of the rigorous terms for instrumental analysis in the next chapter and the geometrical mapping techniques in the one after that.

Motifs or gestural cells impose particular structures on the trajectory through time of a single parameter (or a few parameters, commonly pitch and duration). In the chamber work *COALS* we will see how multiparametric motifs and individual parameter streams can coexist, both structurally and perceptually.

Continuous rendering is analytically simpler than discrete rendering. It leaves more room for imagination, less for precise structure.[58] It applies particularly to timbre, spatial position, time, and stochastics. The concept of timbre space introduced by Grey (1975) fundamentally extended the idea of continuum beyond pitch, amplitude, and time to all possible acoustic parameters by rooting them in perception of sound rather than generation of sound. Spatial distribution of sounds derives its inherent continuity from that of space itself. Even in the case of mechanically separate instruments, spatial distribution is arguably continuous by means of crossfading between adjacent performers. In the absence of pulse, tempo cannot be perceived and time becomes continuous. Boulez has used profusions of grace notes instead of precise rhythmic durations to notate such effectively nonpulsed passing of time (*COALS* plainly borrows this method as well). Stochastics is also fundamentally continuous. Defining a probability distribution for a parameter, static or varying as time passes, is most easily done over a continuous interval, a range of values. If the probability density function is itself parametrically defined, these parameters too can be varied continuously.

Because of limitations of performer or instrument, a parameter might in practice take on only a finite, discrete, subset of the total possible continuum of values. Many mathematical structures can be imposed on this subset, this resulting alphabet of values. Markov processes add discrete probability distributions to the alphabet, allowing discrete motives and fragments thereof to be displayed. Symmetries can be imposed on the alphabet, as with Messiaen's modes of limited transposition or the explicit symmetries encoded in algebraic structures like groups. Sieves and weighted sieves combine the probabilistic aspects of Markov processes with symmetry. The most pervasive structure imposed on an alphabet of values is modular structure, historically based on octave equivalence of pitch.

Serial organization is often coupled with modular structure, though it has been built on simpler structures like additive or divisive series of durations, or even non-orderable "parameters" like attack articulation (though without an ordering, transposition and inversion are difficult to define meaningfully). Its common feature is the arrangement of letters from the alphabet into a fixed order, a series; this series can then be modified to produce other series by means of the classical operations of inversion, *etc*. Serialism thereby generates more variety by transferring compositional emphasis from the values of parameters to the intervals between values—the motions of a dancer instead of the poses.

Serialism orders elements *ex nihilo*: it begins by stipulating that the alphabet be presented in some order. Other organizations of the alphabet, discrete structures like grids and groups built on the alphabet, often do not begin with an ordering. But any such structure eventually needs to be subjected to some ordering or other—a traversal—to be directly representable in the musical surface which flows through time. The choice of traversal can affect the music as powerfully as the choice of structure itself.

Finally, besides continuous and discrete rendering, fuzzy discrete rendering embraces the imprecision of both performance and audition of discrete rendering. Here the composer specifies only ranges, not exact values.

1.6.2 Combining parameters

How can separate parameters be combined? How can a single organizing principle be applied to several parameters? The historical precedent for wanting to do so is the immense effort composers have spent searching for a structure of duration which corresponded closely to their structures of pitch. Arbitrary, separate organization of these two has been annoying to all even if consensus on a single solution has not been reached.

Two parameters which do more than merely coexist, which interact, can equivalently be said to have some constraints applied to them. If the constraints are so tight that the value of one parameter is determined exactly by that of the other, we may do better to speak of one rather than two parameters. Between these extremes of total independence and total agreement, we can play with the degree of correlation between the parameters.

Constraints can be defined with recourse to geometrical analogies, representing individual parameters as orthogonal axes of a space. We have seen how the sphere is a model for continuous rendering and how two- or three-dimensional grids can be used for discrete or fuzzy discrete rendering. Geometrical arrangements of parameters other than orthogonal axes are also possible: they can be concatenated, as in Stockhausen's fusion of the conventionally independent continuums of duration and pitch by identifying short duration and low pitch. More detailed exploration of a region of such a parameter space is possible with space-filling curves like Peano or Hilbert curves, or self-avoiding random walks on a lattice. Not

---

[58] It is odd therefore that Boulez lauds imagination and abhors the continuum.

only do such methods group several parameters together, they define a traversal through the space defined by these parameters, in essence a component of a musical score. Noninstantaneous constraints between parameters introduce time into the discussion. A simple example of this is canons of parameter, analogous to conventional canons of pitch and duration. Such constraints can also be seen as being between parameter streams instead of between parameters, that is, having a temporal component. When defining constraints between parameters, we can also consider how to manage and contain the "combinatorial explosion" of possibilities resulting from this combination of parameters; in other words, we can consider the opposition between redundancy and high density of information.

Intermediate-level structure can be generated from the power set of a set of values (exploring all its subsets). Following Babbitt's example, each subset can be individuated by means of some parameter other than the parameter over which the set itself varies. When the power set is intractably large, as is often the case, reduced collections of subsets can be explored. A convenient collection is one which covers the set without duplication, *i.e.*, partitions the set. Again, the subsets (the elements of the partition) are best individuated by means of a parameter other than the parameter over which the set itself varies. The analysis of *COALS* in the final chapter offers many examples of all these methods. Finally, combining the spirit of power set with the letter of partitioning produces an all-partition array, the presentation of all possible partitions of a fixed size.

The frequency multiplication of Boulez similarly combines two concepts, a pitch class series and a partition, to form a rich structure. Frequency multiplication is atypical in that its combination is rendered again in pitch class, instead of in a different parameter as Babbitt does with partitioning.

Finally, constraints can be resolved by the performer(s) instead of by the composer. This is often done by means of a score which is not a fixed sequence of events, but consists of separate events and rules for combining these events. Resolution of constraints is more difficult in real time, of course, so in this case the constraints should be quite simple and even memorizable.

2. Historical Overview of Instrument Design

We have seen how a building block of software, the function call with parameter list, has been applied to music by composers: spaces of parameters, constraints between parameters defining subsets of these spaces, the combination of separate formalisms to produce richer structures, and so on. Now we turn to another group of musicians, instrument builders (and, necessarily, performers). Much wisdom is found in the design of traditional musical instruments, and an accelerated wisdom is forming about newer instruments. This chapter examines the function of musical instruments from this point of view of nested function calls: the performer puts members of a well-defined family of gestures into the instrument, the instrument then maps these gestures to sounds, and sound then comes out. This theoretical apparatus, like much of Boulez's theory, can be unwieldy for specific instances: why bother with all this to describe the plucking of a guitar string? But its generality provides a language for designing new instruments such as the eviolin described and evaluated in chapter four.

A note about mappings: the connection between input gesture and output sound is quite direct for traditional instruments but it can be arbitrarily intricate for instruments which contain software. This chapter covers only general attributes of such mappings. The next chapter considers in detail various ways to define mappings for new instruments, including a general method which is demonstrated with the eviolin.

Musical instruments have been catalogued and arranged into taxonomies in various ways by scholars over the years. Mahillon (1893) divided instruments into autophones/idiophones, membranophones, aerophones, and chordophones; von Hornbostel and Sachs (1914, 1961) expanded on this. Following Mahillon, instruments are usually grouped by physical structure or means of sound production. Such groupings are inappropriate for *synthetic instruments*, where the connection between the performer's controls and the produced sound is mediated by software.[59] (By *control* we mean a part of an instrument intended to be manipulated by the performer to affect the produced sound.[60]) Instruments found in symphony orchestras and chamber ensembles we call *orchestral instruments*. Ideally this term would describe precisely those instruments which are not synthetic, but the boundary is blurry.

---

[59] A more recent taxonomy (Kvifte 1989) comes closer to dealing with synthetic instruments. It considers the actions of a performer on an instrument's "control organs" (for example, plucked violin and bowed violin are placed in separate families), but it does not address electronic instruments in any depth.

[60] The literature of control theory calls this a *control junction*, "a device by means of which energy is released or regulated as a function of a control signal" (Kelley 1968, 22).

A synthetic instrument is literally the synthesis of a set of controls and a means of producing sound, connected by software. Often the mechanism of sound production is also in software, computing an acoustic signal played over loudspeakers. Such instruments are sometimes called *reconfigurable*, as software is apparently more pliable than hardware; indeed, arbitrary mapping from controls to sound is practical only with software. However, we do not presume such pliability once the instrument has passed from the designer to the performer. Performers find it difficult to adapt to an instrument whose behavior changes substantially during a performance, so we consider an instrument to have a relatively fixed interface and means of producing sound.[61] Ungvary and Vertegaal (1999) also note this difficulty, with what they call cyberinstruments. The term "virtual musical instrument" or VMI is also sometimes used, though this suggests that such instruments are somehow not real.

We assume that an instrument has exactly one human performer: no more, because rigorous analysis of the instrument is problematized by the interaction between performers; no less, because sound sculptures such as wind chimes have no controls for a performer, eliminating any discussion of interface. (One could stretch the point and call the wind or similar phenomenon a performer with unusual physical and mental properties: no ears and no will.) The instrument should also have at least one human listener. More than one does not change our analysis, as the listeners are taken to be absorbers of the instrument's sound and not participants (by way of acting on the performer, instrument, or other listeners; chamber music is a separate case). On the other hand, less than one listener means that we can analyze the sound signal in arbitrary ways not constrained by the human hearing apparatus. This would permit a range of analyses too broad to draw useful conclusions from.[62] Of course there may be only one listener who happens to be the performer. This occurs often enough in private rehearsal, but also for simple enjoyment. But as others can hear nearly the same sound as that which the performer hears, we can generalize the private performance to a slightly less private performance for a few trusted friends.

---

[61] "The mental load of dealing with a constantly changing system will never allow a musician to internalize the system and achieve efficiency and effectiveness" (Vertegaal, Ungvary, and Kieslinger 1996, 308). This approach is also taken by one of the fathers of *musique concrète*, another field where potentially any collection of sounds can happen: "Any device which permits the obtaining of a collection of varied sound objects, all the while maintaining the appearance of an unchanging source, is a musical instrument" (Schaeffer 1966, 51). Much of the rest of this book examines this idea of how we perceive as unified the sounds from a single orchestral instrument, and how much of this can be applied to sound objects (*objets sonores*) divorced from their means of production. Software synthesis of sound shares this danger with *musique concrète* when it transcends the naïve notion of timbre, that which is heard as common to all the sounds of an (orchestral) instrument. Schaeffer (1966, 52) also emphasizes the importance of classifying musical instruments "not any more from the instrument in itself, but in the relation which obtains with the instrumentalist."

[62] If a tree falls in the forest and no one hears it, does it make a sound? This reasoning answers: Yes, but the sound cannot be named.

We need a way to describe an instrument in context with its performer and listener, one systematic enough for both synthetic and orchestral instruments. Again noting the separation of controls from sound production in synthetic instruments, we hypothesize that a general interface between these two can be said to consist of a finite set of (continuous) scalar values and a finite set of discrete values. These values we call *parameters*.[63] Synthetic instruments whose sound is produced by software satisfy this hypothesis, because their sound synthesis algorithm must be controlled by some such set of parameters explicitly encoded in the software.[64] Orchestral instruments can also be described by this hypothesis of finite interface as we shall see. We then have four questions to ask about an instrument:

- What controls does the performer have over the instrument? (What can the performer do?)
- How do these controls affect its sound?
- How does the performer conceptualize these controls?
- How do composers notate the performer's actions? (How do *they* conceptualize these controls?)

After analyzing several instruments, we note common patterns observed. These patterns direct us towards answers for another set of questions:

- How should one design the interface which a synthetic instrument presents to its performer?
- Which physical gestures work well with which musical gestures?
- How many things can a performer pay attention to at the same time?
- How flexible can an instrument be without baffling its performer?

One set of answers is realized in the eviolin.

2.1 A survey of orchestral instruments

In analyzing the common orchestral instruments, we aim to discover patterns of how musical performers play their instruments rather than to present techniques of orchestration. To simplify our analyses, we consider these instruments primarily as their designers imagined them to be played. Most

[63] This is more definition than hypothesis. We exclude basic types other than these, in particular enumerations (nominal rather than quantitative data) and character strings (labels). This does not appear to limit our discussion or exclude basic musical behaviors. This definition does, of course, include arrays and heterogeneous structures of real-valued and integer-valued numbers. Finitude of parameters may not hold in spirit with arbitrarily long arrays; more on this later.

[64] Since this discussion is restricted to instruments with fixed behavior, it also excludes changeable parameter sets or "programmable" instruments. Software design and music performance involve different kinds of gestures and challenge different limits of the human being, to put it mildly.

instruments can be struck, prepared, or played in unorthodox ways to good musical effect, but analyzing such techniques does not lead to significant conclusions beyond those found from observing orchestral instruments as a whole, and the percussion family in particular.

We first define some terms. This dry language cannot afford poetry since it must apply to all musical instruments, not just a few favorites. Once a single instrument is being discussed, poetic language inevitably and appropriately returns to accompany these rigorous concepts.

A *control* is an indivisible part of the instrumental interface, presented to the performer. Its instantaneous state can be described by a finite, often small, set of discrete and/or continuous parameters.

We call the instantaneous state of a control its *value*.

A control is called *discrete* if its state is purely discrete, *continuous* if its state is purely continuous. (In the literature on computer-human interfaces, discrete and continuous controls are sometimes called *buttons* and *valuators* respectively.) Controls with partially discrete and partially continuous state are rare in practice, so we give them no special name.

A *dimension* is a linear continuum.[65] Its *value* is a scalar, the dimension's realization at an instant of time. Often a dimension is associated with a single scalar control, in which case we identify the values of the dimension and the control. Sometimes we speak interchangeably of *parameter*, *dimension*, and *degree of freedom*.

A continuous control *drives* a dimension if a change in its value produces a corresponding change in the dimension's value.

Since one control may drive several dimensions, and again, several controls may drive the same dimension, we can imagine a column of points corresponding to controls, a second column of points corresponding to dimensions, and arrows from points in the first column to points in the second. We call this diagram the *driving graph* of the instrument (figure 10).[66]

[65] For most purposes a dimension is homeomorphic to the unit interval, though in some cases it need be only locally homeomorphic to the real line. (Two spaces are said to be homeomorphic if there exists a continuous bijection between them with continuous inverse; they are then equivalent as topological spaces.)

[66] A graph is a set of points together with a set of edges connecting pairs of these points; graph theory is the branch of mathematics dealing with such structures. Though the bipartite driving graph is often not connected, it never has singleton connected components (if it did, we would not have bothered to name the corresponding nondriving control or nondriven dimension).

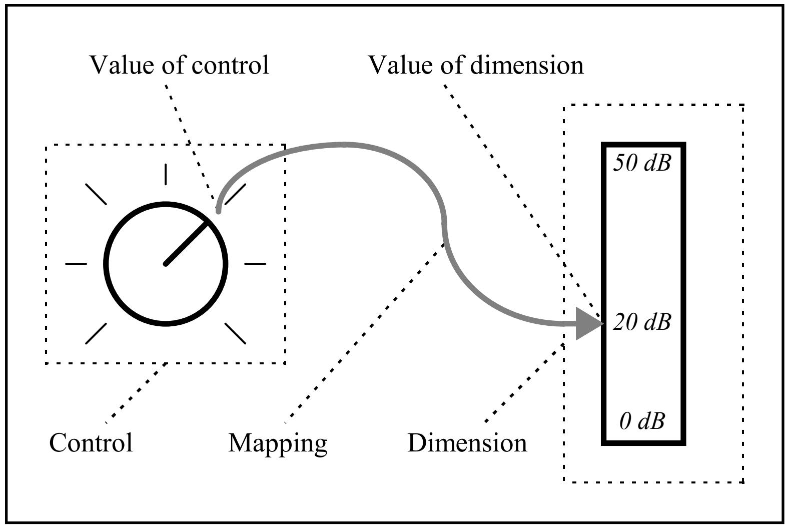


Figure 10. The simplest nontrivial driving graph of an instrument: one scalar control driving one dimension.

A set of controls is *cross-coupled* if, together with the dimensions they drive, the controls are all in one connected component of the driving graph. We investigate this in the context of other mathematical issues in the next chapter.

The sound produced by an instrument changes as the performer adjusts its controls. This changing of the controls' values changes the values of the dimensions driven by these controls. So we represent the sound not as a waveform or as a plot of amplitude versus frequency versus time, but as the change of value of several dimensions with respect to time. (This representation more literally reflects what the instrument turn the printed musical score into.) These dimensions we call *acoustic dimensions*, scalar values computable from conventional representations of the sound. Common examples are fundamental frequency, loudness, and spectral brightness (how much high-frequency energy a sound has, relative to its fundamental frequency). In itself an acoustic dimension is an abstract continuum, any real-valued function of the sound waveform generated up to the present moment; in practice we concern ourselves with acoustic dimensions which are perceptually important or at least perceptible.[67]

A discrete control is also called a *switch*. One of the finite number of states which the switch can assume may be an idle or nil state. A switch in this state is called *disengaged*; otherwise it is *engaged*.

[67] Nonacoustic dimensions are possible. If we let a control drive a dimension d_1 which itself is taken to be a scalar control driving some other dimension d_2, then d_1 is not a direct property of the sound, not acoustic. But this stretches the definition that controls are *directly* manipulated by the performer, and destroys the bipartite property of the driving graph by introducing intermediate nodes (in the manner of a neural network).

A *momentary* switch remains engaged only as long as force is applied to it. A *latched* switch maintains its state without the performer continuing to apply force to it. A *binary* switch has only two states.

We call a continuous control *scalar* if its value is a single scalar.

A continuous control has *hysteresis* if its output depends on its previous inputs as well as its current input, showing some reluctance to change. The control behaves as if its current value were somewhat closer to its recent values than it actually is. In *anti-hysteresis* or overshoot the current value is nudged away from recent values, though such controls are less common.

The *order* of a scalar control (with respect to a driven dimension) is an integer which describes how direct the mapping is between the physical position of the control and the value of the dimension. An order-zero control has a direct relationship between physical position and acoustic dimension (*e.g.*, a volume knob). A control of order one has a direct relationship between physical position and *rate of change* of dimension (*e.g.*, a knob adjusting how fast a pitch rises or falls in a glissando). A control of order −1 has a direct relationship between physical *speed* (rate of change of position) and dimension (*e.g.*, a violin bow driving amplitude).[68] Note that to speak of the order of a scalar control, we presume only that the mapping is continuous; in particular, the mapping need not be linear.[69]

We prefer an alternative notation for the order of a control, borrowed from the field of PID (*proportional-integral-derivative*) controllers (Dean and Wellman 1991, 144). If $x(t)$ is the input physical position of the control at time t and $y(t)$ is its output value, then a *proportional control* can be expressed as $y(t) = \eta(x(t))$ for some mapping $\eta{:}\mathbf{R}\to\mathbf{R}$; an *integral control* has the form $y(t) = \eta(\int_0^t x(t)dt)$; and a *derivative control* has form $y(t) = \eta(\frac{d}{dt}x(t))$. In other words, the value of a proportional control depends purely on its input, that of an integral control depends on the history of its input, and that of a derivative control depends on the rate of change of the input: orders −1, 0, and 1 correspond to derivative, proportional, and integral controls respectively.[70] PID control has proved adequate in industrial processes for a century; orders greater than 1 or less than −1 have been rarely needed. We shall see that they are rare in musical instruments also, so the terms proportional, integral, and derivative control almost always suffice.

---

[68] In a control with order n, acoustic dimension varies as the n^{th} integral of physical position; in a control with order $-n$, acoustic dimension varies as the n^{th} derivative of physical position. Kelley (1968) shows that high-order control systems are typically multi-stage, a 4th-order control driving a 3rd-order, ...driving a 0th-order control.

[69] For reasons of extensibility, this separation of discrete and scalar controls departs from the three-level hierarchy of discrete ("set-and-forget" auxiliary) controls, 0th-order scalar controls, and 1st-order scalar controls presented in (Vertegaal, Ungvary, and Kieslinger 1996, 310). A discrete control cannot be said to have an order, when it is not based on an underlying continuum on which differentiable functions can be defined.

[70] A mnemonic: the numbers −1, 0, and 1 correspond to the change of degree of a polynomial which is differentiated, unmodified, or integrated respectively.

We shall see that several results from PID control theory also apply to musical instruments. Proportional controls are the most common, so for brevity in this chapter a control is taken to be proportional if not stated otherwise. Derivative controls (like a violin bow driving amplitude) are more agile than proportional controls, but fare worse at holding a constant value. Integral controls are, not surprisingly, the opposite: they lack agility but once they reach a desired output value they easily maintain it (by zeroing their input: once $x(t)$ is zero, as t continues to increase then $\int_0^t x(t)dt$ remains constant). Derivative and integral controls can also be characterized in terms of their transient response, the ability to execute sudden changes in value. Derivative controls have good transient response while integral controls, especially ones with large gain, tend to overshoot the desired output value. Finally, integral controls work better when the controlled system has high hysteresis (Dean and Wellman 1991, 148–149).

The *gain* of a scalar control (again, with respect to a driven acoustic dimension) describes how strongly input affects output. For the same range of input values, a high-gain control has a wider range of output values while one with low gain is easier to play. A control may have variable gain.[71] In terms of the preceding paragraph, gain can be thought of as a scaling transformation applied to the mapping η. A control with zero gain has an unchanging output, and can thus be excluded from a driving graph.

A dimension is *fretted* (with respect to a control driving it) if it takes on only a finite set of values from its potential continuum of values. Since human hearing has limited resolution, this definition has a perceptual condition: all members of this finite set should be distinguishable by a listener (the values should be perceptually separable).[72] For many orchestral instruments, frequency is fretted with respect to their primary frequency control (keys or valves, or a fretted fingerboard). The difference between a switch and a fretted scalar control is that only the latter's states have a natural ordering induced by the continuum of the dimension. A dimension may be fretted with respect to some controls and not with respect to others, as shown by the frequency of a guitar string, driven by (i) finger position and (ii) transverse pulling of the string.

A continuous control may have one or more *detents*, positions in its range of motion where motion is hindered. Detents are to frets what suggestions are to commandments: if the performer pays little attention to the control, it will tend to assume the value of one or another detent, but with more attention the control can assume values in between. *Basins of attraction* of detents can have different sizes. If

[71] For example, a fingered violin string has higher gain farther up the fingerboard. Near the top of the fingerboard, moving the finger a millimeter greatly changes the pitch, which places more notes under the hand but makes playing in tune more difficult.

every position of the control lies in some basin of attraction, then the control is guaranteed to tend to a detent value if the performer releases it (*e.g.*, a pitch-bend wheel or a self-centering joystick, where the control space is a single basin of attraction).

A control has a *resolution*, its measurable accuracy. When one scalar control drives one dimension, we may similarly speak of the resolution of the dimension.

The *temporal resolution* of a control describes how quickly the performer can set the value of the control within some margin of error. Most generally, this is a function defining duration as a function of admissible error, starting value, and ending value. This is often simplified by collapsing starting and ending value into distance or by holding the amount of error constant.

A dimension may have particular *gestures* which are performable in it, particularly if it is driven by more than one control. Violin left-hand technique drives pitch with several controls: hand position, finger placement, rate and amplitude of vibrato. A gesture, a perceptually identifiable grouping of control changes, can be said to occur when several of these controls are changed at once. One example of this is the sequence: while stopping a string with the index finger, slide into a higher position; stop a new note with the little finger; hold the new note for a moment; resume vibrato on the new note. Another example from the final movement of Beethoven's Violin Concerto is the alternation of successive pitches between an open-string pitch and a melody played higher on the same string. Gestures can also be performed with a control driving several dimensions, as easily as with a control driving one dimension.

Finally we note that, like mathematics itself, these definitions are built on undefined objects. We can contain the problematic aspects of one undefinable term, parameter, by artificially dividing it into *inherent* and *noninherent* parameters. Any acoustical parameter space has the former, those defining it, and the latter, those added by the performer. Instrumental performance practice illustrates this: tremolo belongs to the violinist, not the violin; to the piano the pianist adds relative loudness of simultaneous notes as well as amount of legato from the temporal overlap and similar loudness of consecutive notes of nearby pitch. All instrumentalists may add the two parameters strength of downbeat and evenness of rhythm. The list can go on to arbitrary levels of detail. Fortunately we can take a cue from cybernetics and analyze this situation by considering a new system consisting of instrument plus performer; this combination is reasonable because orchestral instruments developed in response to performers, not in isolation from them.

---

[72] Without this constraint, software-based synthesizers would be fretted in *all* their acoustic dimensions and the term would lose meaning.

We have (i) defined the controls and acoustic dimensions of an instrument, (ii) presented some categories and possible properties of each, and (iii) presented possible properties of the relationships (mappings) between controls and dimensions. Now we analyze the common orchestral instruments in these terms.

2.1.1 Woodwinds

Woodwinds (flutes, oboes, clarinets, bassoons, saxophones) are controlled by the diaphragm, mouth, and fingers. An array of fingered momentary switches, in combination with mouth shape, drives frequency.[73] The frequency dimension is fretted with respect to these discrete switches, but not with respect to mouth shape: the latter adjusts frequency slightly above or below a fretted value. Mouth shape also drives frequency in a coarse manner, changing in which overtone region the instrument resonates. This last driving mechanism has significant hysteresis, since the instrument is reluctant to change its mode of resonance.

The diaphragm along with chest and cheeks drives amplitude and spectral brightness. This is a derivative control, since rate of contraction, not position, of these muscles produces airflow and flow rate (volume per unit time) is proportional to amplitude. The precise function defining airflow in terms of rate of contraction varies considerably among players; even the function of a single individual depends on fitness level, as studies of athletes have shown. With certain instruments circular breathing is possible (inhaling through the nose while continuing to provide outward airflow from only the cheeks); this gives the illusion of unlimited diaphragm travel. The bagpipe allows a kind of circular breathing by introducing an air reservoir between the player's mouth and the sounding pipes. Blowing into the reservoir's blowpipe and compressing the reservoir with the arm combine to drive amplitude, although pitch wavers as a side effect so pipers try to play with constant amplitude.

Tongue tension just before the onset of a note (forcefulness of a 'T', 'K', or glottal attack) drives the spectral brightness of the attack. Lip tension takes this role, when lips rather than front or back of tongue hold back airflow from the lungs to the instrument. Again, the physical phenomenon at work is how the airflow varies with time. Greater muscle tension produces a greater initial transient airflow, resulting in more high-frequency sound. The greater the variation of airflow before it stabilizes after a few tens of milliseconds, the greater the transient variation of frequency, amplitude, brightness, inharmonicity,

[73] In this section on instruments, fundamental frequency is assumed by the generic term frequency. We avoid the perceptual term pitch for the moment.

relative amplitude of noise, and other timbral parameters. Once the tone reaches its steady state, mouth shape drives other aspects of timbre.

Once fingering has been drilled, the attention of the player is mostly on intonation, steady airflow, "good tone." In fast passages, or between two notes whose fingering change involves the simultaneous engaging and disengaging of many switches, attention reverts to fingering.

2.1.2 Brass

Brass instruments (trumpets, horns, trombones, tubas) are quite similar to woodwinds. Having historically developed from valveless designs, they depend less on fingering: a smaller array of fingered momentary switches together with mouth shape (primarily lip tension) drives frequency. We disregard half-valving and consider the switches to be binary. As the number of overtone regions is considerably larger than the four or so available to woodwinds, we cannot simply separate mouth shape into fine and coarse frequency controls (witness the difficulty of playing the horn in its highest register). Lip tension then drives frequency in a not quite fretted way. With ideal fretting, the function mapping lip tension to frequency would be a step function, in this case a "staircase," but in practice each corner of the staircase is rounded in a hysteresis curve. Novice players try to avoid these rounded corners and keep lip tension in the middle of each step, but experts can play accurately on the rounded corners to bend pitch precisely.

Diaphragm, tongue/lip tension, and mouth shape affect timbral parameters as with woodwinds.

Instead of an array of valves the slide trombone has a slide whose extension drives frequency. Some trombones have one or two valves as well which extend the lower range. Slides have been occasionally given to other brass and woodwinds, for instance the slide saxophone. Placing the slide in different positions produces several overlapping harmonic series. This often results in several possible slide extensions / overtones for a given frequency. Choosing the best one is rather like choosing a fingering on the piano. But once a choice has been made, the attention of the performer is directed much like with woodwinds (though a little more is directed towards intonation).

Unlike woodwinds, most of the sound of a brass instrument radiates from a single point. This allows brass instruments to be easily muted by obstructing the bell. Some mutes are binary switches (latched: Harmon, or momentary: cup) affecting timbre and amplitude, typically attenuating a particular frequency band and concomitantly reducing amplitude when engaged. Other mutes are better considered continuous controls, similarly driving timbre as their distance from the bell varies. Some mutes, notably "stopping" the horn with the hand, drive frequency as well (continuously, as the distance from the bell varies); they effectively change the length of the vibrating air column.[74]

All orchestral instruments radiate sound nonuniformly, especially the brass family: amplitude increases suddenly as the listener approaches the main axis of the instrument, which has led to performance indications like "bells up." The radiation patterns of other instruments, notably strings, do not admit to such easy general characterizations; at best we can say that orientation of the instrument relative to the listener drives timbre (frequency attenuation and amplification) as a complicated proportional control. Besides spatial orientation, two other spatial properties also affect what the listener hears, distance and relative velocity between instrument and listener. These two drive amplitude, amount of reverberation, and Doppler shift. But as these properties are common to all sounding bodies, behave identically for all, and are difficult to adjust quickly during a performance (*i.e.*, use as compositional parameters), we do not treat them here.

2.1.3 Voice

Speaking and singing are similar to brass and woodwinds in two respects. First, the diaphragm is a derivative control of amplitude and spectral brightness (and frequency, slightly). Second, mouth shape drives steady-state timbre, called vowel color in this context. Position of tongue, lips, and jaw (*i.e.*, teeth) drives the unpitched noisy component of the sound (consonant "color") in overall amplitude and in spectral content; when an unsustained consonant starts from silence, the immediately preceding tension in tongue and lips also drives consonant color in this way. Vocal cord tension drives pitch, with hysteresis for small changes and sometimes anti-hysteresis for large leaps. Vowel color can be described in terms of spectral peaks called formants; five formants suffice to distinguish most vowels, though two or three often suffice. As with brass, continuous muting is possible though rare; one method is cupping hands in front of the mouth. Some choristers bend one ear towards their mouth with an extended hand, as a trick to hear themselves better in an ensemble. (Insecure choristers sing into their music folders, too.) All such mutes drive only timbre and amplitude as their distance from the mouth varies; pitch is unaffected.

An anatomical atlas such as (Gest and Schlesinger 1995) reveals that mouth shape (again, tongue, lips and jaw) is high-dimensional, though some research indicates that the effective cognitive dimension is

[74] The harmonica can be considered a compound woodwind: it is small enough to be covered by the hands, and this muting is important in certain popular styles. Also, *any* instrument can be "muted" by playing it offstage; this blurs the definition of instrument to include the concert hall, though, and opens the question of how else the hall can then be controlled.

somewhat lower (Maeda and Honda 1993).[75] For any instrument, as someone masters it and directs proportionally less attention to low-level muscular action, the dimensionality of their mental model of the instrument—the size of the state space which the performer-plus-instrument inhabits—likely becomes smaller in the same way. Experiments with chess players memorizing board configurations found that while novices did equally well on any kind of configuration, random or legal, good players fared much better with legal ones. Masters most easily memorized board configurations taken from games of good players (Chase and Simon 1973).

2.1.4 Strings

We first discuss left-hand pitch technique for stringed instruments, and then deal with right-hand string excitation: plucking and bowing. Two special cases, electric amplification and quasi-strings, round out our discussion.

If only one finger stops a string, then string length and correspondingly amount of extension of the left arm drives frequency (proportionally, but with variable gain). This is of course oversimplistic: the actual multiplicity of fingers leads to a compound analysis. Arm extension drives frequency coarsely (what "position" the left hand is in). Within the range there available to the fingers, a five-state momentary switch, finely driving frequency, is described by which finger stops the string (four stopping fingers, plus the open string). Yet again we need to refine the model: on the violin, a range of nine semitones, not just four fingers, fits under the hand, and microtonal playing extends the states of this switch past 9+1 to even higher numbers. If the entire pitch continuum is used, we describe the action of the left hand as a separate frequency control, fine-grained as distinguished from the coarse control of the left arm. If vibrato is used, rotational position of the fingertip along the string is a last extra-fine control of frequency.

The right hand may also adjust frequency finely by moving the bridge, directly as with some Eastern instruments or indirectly as with the whammy bar of an electric guitar; this last only decreases frequency, and has a detent at maximum frequency.

If the fingerboard has frets, different controls drive frequency at the fine level. Instead of raising and lowering frequency by rolling the fingertip along the string, frequency can be raised (not lowered) by

[75] Seven independent muscles work the mouth, the jaw has seven more, and the tongue another eight. Simplistically considering these to work in pairs like most flexors and extensors, mouth shape is a subset of a space with $(7 + 7 + 8) \div 2 = 11$ dimensions.

deflecting the string transversely where it contacts the fingerboard, thereby increasing its tension.[76] This is easier when the string has lower tension, typically for long strings. A fretted fingerboard also simplifies finding a particular pitch by reducing the number of pitches to choose from. As a corollary it also facilitates double-stopping.

With plucked strings left-hand control is similar to that of bowed strings, but since amplitude is not sustained (particularly on short strings) vibrato and portamento are correspondingly less prominent.

Classical guitar plucking consists of excitation, release, and damping (Cuzzucoli and Lombardo 1997). Excitation starts with a displacement of the string of a few millimeters, either gradual (*tirando*) or sudden (*apoyando*); this may coincide with and affect the damping of a previously sounding note. A gradual instead of a sudden release, sliding the string past the fingertip, attenuates the higher partials of the resulting note. Damping can similarly be gradual or sudden. So amplitude is driven primarily by finger and forearm tension just before a note begins (recall the 'T' of a woodwind attack). Spectral brightness is driven by distance of the plucking position from the bridge and also by finger tension in the brief moment during release. Rate of decay during the steady-state portion of a note is uncontrolled, but during damping it is driven by tension of the damping finger or hand.

If the hand holds a plectrum, rapid up-down alternation is possible. This simplifies playing fast runs of notes but makes crossing between strings more awkward. Plucking directly with the fingers, only flexion not extension is practical; alternating between fingers is one workaround. But unlike with the plectrum, bare fingers allow several strings to be plucked simultaneously. Plucking several strings at once with the fingers or nearly at once with a stroke of the plectrum, *i.e.*, strumming, traditionally plays an accompanimental role of harmonic support for a separate melody.

Tremolando repetition of a single pitch is commonly heard and notated as a single note, as with the mandolin. It allows for coarse effects of crescendo, decrescendo, and controlled attack and decay. We describe it as a hybrid of plucking and bowing. Limited timbral control derived from that of plucking applies, while average forearm tension and rate of repetition combine to drive amplitude. Rate of repetition clearly affects spectral brightness: more attacks per second implies more bright attack transients.

When a string is bowed, extension of the bow arm acts as a derivative control of amplitude. The distance along a string from the contact point of the bow to the bridge drives spectral brightness (*sul tasto*, *sul ponticello*); downforce of the bow on the string also drives spectral brightness (light "skating"

[76] Such transverse bending is of course also possible on nonfretted strings, but offers nothing beyond conventional change

attenuates the fundamental and emphasizes high-frequency noise). Arm extension and bow force are controlled by a complicated interplay of shoulder and elbow rotation, axial forearm rotation, and finger contraction; the purely muscular dimensionality of bowing, like that of vowel color in singing, resists formal analysis.

The modern Tourte bow can interact with the string in many ways. Nearly inaudible bow changes, by reducing pressure and letting the string resonate by itself for the briefest moment, give the illusion of an arbitrarily long bow. Subtle distinctions of changing bow force through a stroke are expressed by the terms *détaché* and *spiccato* (separated, picked), *martelé* (hammered), *sautillé* (hopping), *ricochet* (rebounding), *saltando* (jumping), *staccato* and a dozen others.

As with the array of keys on a clavier, most bowed string instruments have an array of strings sorted by frequency.[77] With varying agility depending on bridge curvature and bow hair tension, multiple strings can be bowed and fingered simultaneously.

Electric stringed instruments, notably electric guitar but extending throughout the stringed family, augment or replace the physical resonator with a signal pickup and electric amplification. Additional controls are provided for affecting the electrical signal in its electrical (nonacoustic) form. These include a volume knob directly driving amplitude, and often several latched switches or knobs affecting timbre. These switches and continuous controls are often within reach of the right hand or the feet. They may select different pickups, apply filters, or modify the sound in ways peculiar to the electric domain: moving the signal between several loudspeakers, adding delays, pitch shifts and reverberation, or triggering auxiliary sounds at certain times.

Feedback, most familiar with guitars, lies in both the acoustic and the electric domain. It is simply described by a single proportional control: amplitude of feedback signal is driven by distance of the instrument body from the loudspeaker. Other factors like orientation of the instrument and other sonically absorbent or reflective bodies near instrument body or speaker cabinet are important but difficult to formalize.

In general, string players direct awareness coarsely towards left hand position along the fingerboard and right hand selection of strings (which ones are bowed or plucked). Finely, left hand intonation requires considerable conscious thought on unfretted instruments, though with bowed unfretted instruments bow technique is responsible for all aspects of sound except pitch and thereby captures even more attention. (Bowed fretted instruments are rare.) With plucked fretted instruments, a fine level of

---

of string length. Even as a redundant control it has not found use.

awareness is directed more towards shape of the hand in chordal passages, and more towards the excitation / release / damping of sequential notes in melodic passages.

Strings are nonrigid bodies which can vibrate when held under tension. Rigid bodies can vibrate by themselves; when they are long, free at one end, and fixed to a resonator at the other, they can be plucked to produce a pitch as in the case of the *mbira* (thumb piano); we call these bodies *quasi-strings*. The term idiophone does not apply here, since only a part of the instrument is vibrating to produce a particular sound, not the instrument as a whole. Excitation, release and damping is similar to that of a string except that the plucking position is restricted to the free end of the quasi-string, since the middle is not easily reached if several quasi-strings are mounted side by side.

A quasi-string cannot be stopped to raise its pitch, since this would disable either the free-end exciting mechanism or the fixed-end resonating mechanism. Changing the pitch of a quasi-string could be accomplished by changing its mass (adding weights) or its length (perhaps a sliding point of contact with the resonator like the schoolboy trick of buzzing a ruler on the edge of a desk). Sliding a mass along the quasi-string would effect this continuously. But it is simpler to just provide an array of quasi-strings of different pitch.

An array of quasi-strings, instead of being sorted by pitch, may be presented with the lowest pitch central and consecutively higher pitches alternating left-right outwards. In the *mbira* this is for convenience of the reach of the two thumbs. This arrangement makes smooth scales and glissandos difficult; two-thumb technique also precludes chordal playing.

2.1.5 Percussion

A *trigger* is a momentary input to the system, often with several continuous parameters. It is equivalent to a momentary switch whose state transition is detected by the instrument, plus a set of scalar controls evaluated only when the state transition occurs. So the concept of trigger does not extend our theory, but in practice it simplifies discussion and implementation.

The popular Western drum set, a collection of cymbals, toms, *etc*., can be described as a set of triggers. Each trigger has its own particular set of continuous parameters. Common parameters are position of impact (for circular vibrating bodies, distance from center suffices), impact velocity, mallet mass, and mallet elasticity. The impact of a compound mallet like a brush or Blastick can be considered as a collection of nearly identical simple mallet attacks.

---

[77] This simplifies the playing of scales and arpeggios. Music not dependent on such gestures could benefit from other

Some triggers have an associated scalar control, amount of damping. A vibrating surface is usually damped by pushing with varying force against it with a nonresonant object such as a mallet or the palm of the hand. The hi-hat, however, damps a cymbal with a quite resonant object, another cymbal. This second cymbal also acts as mallet. A drum head can also be damped by resting a cymbal on it; this is best described as a single resonator with two different families of resonant modes.

Some drums (timpani, "talking drum," *dumbek*) have a scalar control which drives pitch (even during a single note). Timpani have pedals which adjust head tension, while the talking drum is squeezed on the side to produce the same effect; the non-striking hand changes the resonant pitch of the *dumbek* when inserted into the bottom of the drum.

As with guitar string excitation, to be thorough the entire impulse which transfers energy to the vibrating body should be modelled. Strictly speaking this may be several continuous parameters changing over the duration of the impulse (a hand drum like the *tabla*); this can be approximated with piecewise linear envelopes to keep the number of dimensions low.

A few instruments traditionally placed in the percussion section of the orchestra are not struck but excited continuously, for example bowed vibraphone, lion's roar, and the rubbing of a superball mallet-head on a drum head. This mode of excitation is similar to that of bowed strings, though the body under excitation is two-dimensional. The more massive the bowed body, the less responsive it is to bowing transients.

Rattles lie midway between continuous excitation and discrete mallet strikes. The control of individual physical collisions is given up in exchange for a fairly uniform control of many collisions, at a very fast or somewhat slower time scale (guiro or Flexatone). The Flexatone, though held like a rattle, can be considered as a conventional mallet instrument with (i) an additional loose linkage between the hand and the mallets, and (ii) a pitch range whose width varies approximately with speed of shaking.

The percussionist's attention moves fluidly between different levels, as different pieces often have quite different collections of instruments. These levels include which mallets are held, the location and size of struck objects, and the choreography of where one's body is and where it is moving to. Individual strokes and rhythms are generally drilled to the point where little attention is needed there, as is the case with fingering on other instruments.

---

orderings such as exchanging the middle two strings on a violin.

2.1.6 Claviers

Claviers (here piano, harpsichord and organ) are polyphonic instruments. The main control of a clavier is an array of fingered triggers called keys; the array is called a keyboard. (Some claviers have pedal keyboards as well.) When a key is depressed, a tone of a particular fundamental frequency sounds; when the key is released, the tone stops. The amplitude of a tone can decay quickly, slowly, or not at all; rate of decay is not really adjustable by the performer. Amplitude of a note can be specified on the piano: striking the key with greater force produces a correspondingly greater amplitude. In muscular terms, during the striking of a key, hand or finger flexors, arm extensors, or torso flexors act as derivative controls of amplitude. During this striking, the key and its action mechanism convert force into sonic amplitude. Unlike the simultaneous layers of muscle control in violin fingering, these layers from torso to fingers are generally engaged one at a time. Which layer is engaged depends on playing techniques such as finger trills, wrist staccato, and fast chordal passages.

The keys of a clavier are sorted by frequency, usually in the common seven white key, five black key arrangement. Most variants (quarter-tone, split black keys for enharmonic equivalents, duplicate keyboard tuned a semitone away for easier alternate fingerings) do not significantly change the range of performable gestures. Radically different gestures—combinations of keys reachable from one hand position—could be played if the keyboard were organized differently, say all the C's at the bottom, then all the C-sharps, and so on. This is the case with the lowest notes of certain Renaissance organs. To conserve expensive metal used in the large pipes in the lowest octave, pitches rarely used for the bass of a vertical sonority (F sharp and A flat particularly) have their keys assigned to commoner pitches like C and G in the next lower octave. Note that simply increasing the number of keys reachable from one hand position necessarily involves reducing the size of each individual key and hence increases the attention required of the performer. The range of gestures is expanded, at the cost of accuracy and speed.

Additional controls may be present which affect timbre: *una corda* pedal, lute stop, various organ stops, couplers between keyboards. These are generally latched binary switches. Some organs have presets (momentary binary switches) which assign values to some subset of stops and couplers. Quite different in character are controls which modify the behavior of the instrument and add resonance. On the piano, (sets of) keys may be depressed silently to enable selective resonances; the damper pedal enables all resonances. With many performance styles, the damper pedal is simply a momentary binary switch. But for some styles like late romantic repertoire, half-pedaling and other subtleties require us to describe the damper pedal as a scalar control driving amount of resonance from zero to full along a complicated path.[78] Finally, harpsichord and organ may have swell boxes, louvers or shutters which drive amplitude, and spectral brightness to some extent, as proportional controls. (This contrasts with the more common derivative controls of amplitude.)

After much training, keyboard players have an intuitive knowledge of chord shapes, fingerings for common figurations like arpeggios and scales, and hand stretches for different intervals. Only when these vary rapidly and continuously do they command attention; attention is more usually directed towards where the hands are on the keyboard. In music with many leaps such as the later Romantic repertoire, the performer feels several positions simultaneously with one hand. In highly polyphonic music, and in sustaining claviers like the organ, there is a heightened awareness of individual fingers being down or up, depressing a key versus resting on one. Attention jumps quickly among different levels: shape of phrase, fingering, memory (when memorization is required), muscle tension, arm weight, and so on.

2.2 A survey of synthetic instruments

We now describe and analyze several kinds of synthetic instruments. Encyclopedic coverage of this rapidly developing field is impractical, so the particular instruments presented here have been chosen to best illustrate the conclusions drawn at the end of this chapter.

Note that we restrict our analyses to structural considerations, and avoid details of implementation for reasons of brevity. It is unlikely that the reader will fully comprehend an instrument mentioned in this section without examining the references provided for it. This is intentional: we want to define only the controls which the instrument presents to the performer and the sounds it presents to the listener, in general terms which can be applied to the design of new instruments. Every finished instrument has some aspects which transfer to the design of others, and some which are peculiar to itself; we can deal only with the former.

Some synthetic instruments have more internal state at the acoustic level than the performer can keep track of, such as the Sal-Mar Construction (Franco 1974; Martirano 1998) and ImprovisationBuilder (Walker 1994). Detailed acoustic control is necessarily relinquished for high-level algorithmic control, given the finite attentive capacity of the performer. We confine our attention instead to instruments without such internal state, where the produced sound depends almost entirely on the current gesture

[78] Rather than describing this path in terms of which dampers rise and fall first, something unpredictable from one piano to another and perhaps unpredictable from moment to moment on one piano, it is better described in terms of what the performer hears, a feedback loop.

input to the instrument and negligibly on past gestures. This overstates things a little: engaging a secondary control like a violin mute obviously changes the output sound for quite some time. But it does so with high predictability and with no further temporal variation. For example, after three minutes the mute does not of itself start wiggling on and off or decide to retune the strings. The spirit intended here is that the instrument converts gestures into sounds in a highly predictable way; it is not something to have a conversation with, not something designed to surprise the performer.

2.2.1 Emulated orchestral instruments

A synthetic instrument built to emulate an existing orchestral instrument attracts players with the promise of little new technique to learn. Particularly since the mid-1980's this has become commercially viable with MIDI (Music Instrument Digital Interface), a protocol developed by manufacturers of electronic music devices (MIDI Manufacturers Association 1996). MIDI prescribes a format for sending commands from general-purpose controller units to synthesis units. Many general-purpose controllers and general-purpose synthesizers have been built; here we focus on "preset" units or preset behavior within more general-purpose units. We consider the interface presented by a unit to its performer, not to its user or engineer. (Of course the performer and engineer may be the same person. But for the duration of the performance, the persona of the engineer disappears.)

Common MIDI commands are *note on* (start playing a note with this pitch, with this volume, on this *channel* or timbre); *note off*; *pitch bend* (microtonally change the pitch of this instrument); adjustments to the timbre of an instrument like *aftertouch* (finger pressure on a key after it is struck) and other *continuous controllers* (scalar controls, in our terminology); and various initialization instructions to assign particular timbres to particular channels. Most currently available synthesizers are what the industry calls polyphonic: they can play several notes at once on a channel, imitating clavier-based orchestral instruments where depressing several keys produces several pitches, not just (the highest) one.

2.2.1.1 Keyboard controllers

Like pipe organs, MIDI claviers are almost always general-purpose controllers with a large variety of timbres which they can generate.

The keys of a *multiply touch sensitive* clavier sense more than just the velocity with which they are struck. The term seems to have originated with Robert Moog (1982). The best known of these is the Big Briar keyboard controller. Individual keys send: note-on and note-off velocity, vertical displacement, downward pressure, and *x-y* position of the fingertip on the key (Roads 1987; Moog 1987). An earlier example is the Notebender (Allen 2000), whose keys can move forwards and backwards with a single detent. Its two parameters of key displacement and finger pressure (aftertouch) were suggested by the manufacturer to be mapped to individual MIDI parameters such as continuous or nearest-semitone pitchbend, or particular synthesis parameters.

Nonstandard layouts of keys have also been designed, notably for playing microtonal music (the term encompassing any tuning other than familiar twelve-semitone-per-octave temperaments). The historical predecessor of many of these is the Bosanquet Generalized Keyboard equidistant hexagonal layout of 53-tone equal temperament (Bosanquet 1875). Keislar (1987) remarks that its layout preserves hand shape when a passage is transposed.[79] This can be seen in a simpler hexagonal layout of 17-tone temperament (figure 11). Keislar also refers to dozens of other microtonal keyboards and enumerates various hexagonal, rectangular, and diamond-shaped key-grids based on the symmetries of the microtonality the inventor is interested in.

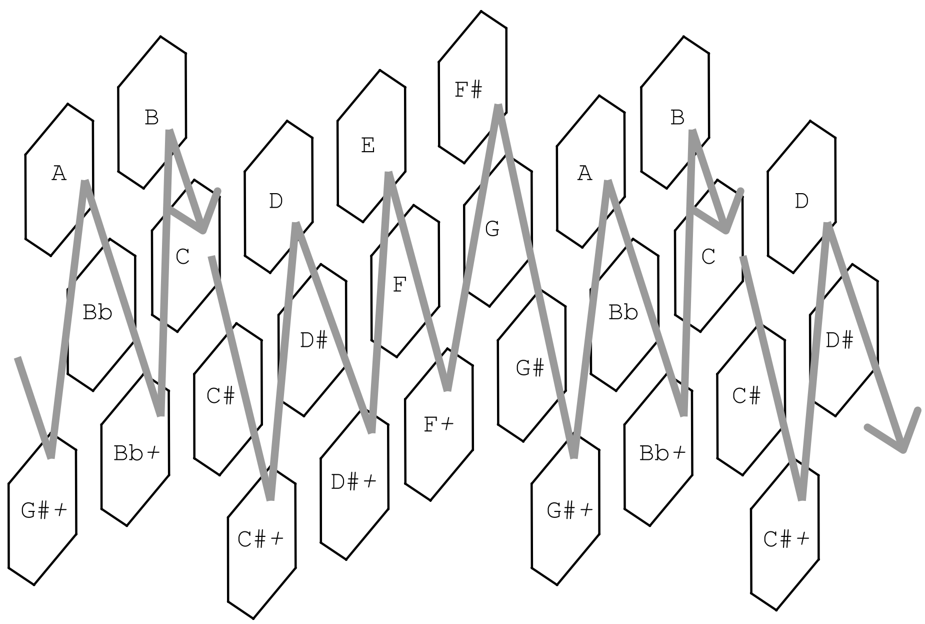


Figure 11. Hexagonal-grid keyboard layout for seventeen-tone temperament. Arrows indicate rising pitch in one octave. Pitches only approximate equal twelve-tone temperament; the notation G#+ indicates a pitch between G# and A. After a diagram by Wilson (1983, 16).

[79] This is approximately true of string instruments, too.

Motorola's Scalatron uses a hexagonal grid of 240 keys (Secor 1975). The more recent MicroZone from Starr Labs has an even larger 9-by-90 hexagonal grid derived from Bosanquet's layout. The keys of the MicroZone lie in a flat plane so glissandi are possible in all directions. Individual keys measure attack force and finger pressure. Individual voices and other parameters can be assigned to arbitrary groups of keys called zones and layers. The keyboard is augmented with continuous controls, a joystick and four sliders. The joystick is called *four-way*: moving it away from center in any of the four cardinal directions generates an increase in one particular parameter (Starr Labs 1998a). A more conventional though still large layout is found in the Monolith, a bank of fifteen four-octave piano keyboards. Within each keyboard, white and black keys are arranged beside each other in a single row (Duringer 1996).

Keislar hints at the compromises of separating out secondary controls, a concept formalized in the next chapter:

> "[T]he key layout can be simplified, at the expense of freedom of selection, with supplementary performance controls for reassigning the pitches of the keys, or with automatic assignment. Control of pitch can be removed entirely from the performer—for example, if the composer's note list is stored in memory. Max Mathews has introduced the notion of a keyboard with only ten keys—one for each finger—for this performance mode." (Keislar 1987, 26)

The larger the keyboard, the more hand position commands the performer's attention. This is particularly so for planar (array) keyboards, where the hands might rotate and translate in two directions, instead of merely translating along one pitch axis. At the other extreme, hand position vanishes from attention for small keyboards like the one Mathews proposes above.

2.2.1.2 Guitar controllers

The first guitar controllers from the early 1980's used piezoelectric pickups on the individual strings of a conventional electric guitar (similar to violin controllers, discussed below). Signals from the pickups were sent through pitch- and amplitude-tracking hardware which drove external MIDI synthesizers. Their slow inaccurate pitch tracking made these controllers unpopular with guitarists. This pitch tracking problem is inherently difficult because of the low range of the guitar. We therefore restrict our attention to "fretwired" guitar controllers which do not track the pitch of vibrating strings, but sense where fingers are pushing against the fingerboard.

The SynthAxe (figure 12) is an early fretwired guitar-like MIDI controller with the addition of right hand keys, a control console and a pedalboard. In addition to scanning the fingerboard to determine pitches, string bend sensors measure how far to raise each pitch. After starting a note with the "trigger key" for that string, the fretted string can be released but the note will sustain until the key is released.

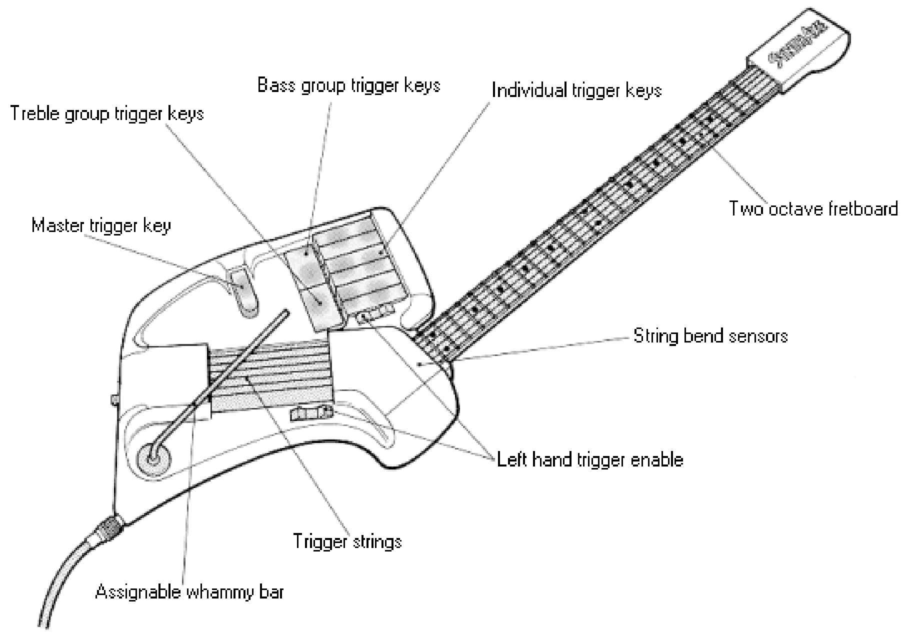


Figure 12. The SynthAxe guitar controller. Reproduced with permission from (Hollis 1998).

The Ztar MIDI guitar controller by Starr Labs senses left-hand finger position with momentary switches in its stringless fingerboard. As with the SynthAxe, the right hand triggers one of six "strings" by plucking one of six very short strings or depressing one of six keys (the latter also measure pressure while depressed). Keys paired with strings offer two families of excitation gestures. For example, tremolando and muting with the palm of the right hand are possible with the strings; independent aftertouch and more independent strumming and plucking patterns are possible with the keys.

When in a mode that plays all depressed frets, allowing multiple notes per "string," the fingerboard behaves more like a planar 6×24-key keyboard, playable with both hands (Starr Labs 1998b). Arbitrary "tunings" are also possible, for instance tuning adjacent strings an octave apart.

The fingerboard can be divided into many (overlapping) rectangular zones of behavior. Behavior here means things like tunings, transpositions, timbre, and loudness. A "solo" toggle lets one zone temporarily take over the whole fingerboard.

A more advanced model of the Ztar is augmented with many secondary controls, in principle assignable to any MIDI function; here we cite typical uses. A dozen switches on the instrument body, under the right hand, trigger percussion. Beside them, a pair of buttons act as an octave switch. Continuous control of timbre is provided by a joystick and a sensor strip along the side of the fingerboard, the latter for the thumb to control a top melody voice. A breath controller drives loudness of a second voice doubling the top melody, and of course pedals can be added too (Starr Labs 1998b).

2.2.1.3 Electronic wind instruments

Yamaha's WX-7 is a well-thought-out electronic wind instrument (EWI) with a profusion of controls, most of which drive pitch. A standard 14-key Boehm woodwind layout plus a multi-octave switch yields a seven-octave range. Two general-purpose trill keys, when depressed, change to the pitch one or two semitones above the fingered pitch. This elegantly demonstrates redundancy of interface making an instrument easier to play. A pitch-bend wheel and/or lower lip pressure sensing adjusts pitch away from its equal-tempered detents. Breath force drives volume as well as user-programmable timbral parameters. As a concession to popular styles a key-hold button transfers the currently played note to a separate (bass) synthesizer, to be sustained there while the performer goes on to play other notes (Blevins 1987).

Compared to orchestral woodwinds, less attention is paid by the performer to good tone and intonation. This is because poor tone is generally impossible with MIDI, and because inaccurate intonation must be intentional. More attention is therefore directed to breath support and fingering.

2.2.1.4 Percussion controllers

The Zendrum has 24 striking surfaces or *pads* played by the fingers. Each pad drives one to four MIDI notes depending on how hard it is struck. This lets the sound of each pad be adjusted by crossfading sounds, in essence driving four amplitudes from one impact force control. It could almost be called a keyboard controller optimized for assignment of unpitched sounds to keys, and for fine timbral control of these unpitched sounds (Zendrum 2001).

Instead of having physically separate pads, some percussion controllers have a single surface which senses the impact position of a mallet. This surface is divided into logical *zones*. Individual zones then have their own behaviors, only incidentally sharing the same physical surface. One example of this is the trapKat, a 24-zone controller emulating a conventional trap-set layout. The trapKat also contains preset rhythmic and melodic patterns which can be modulated during performance (Alternate Mode, Inc. 2000). The drumKat is a similar ten-zone controller with greater programming versatility. Several sounds can be assigned to a zone, each sound with its own velocity response curve (function from impact force to loudness), delay, and gate time (minimum duration of a note). Impact force can also be mapped to pitch and gate time with programmable response curves. A zone can cycle through several sounds as it is repeatedly struck. A collection of zones thus programmed is called a kit; a footswitch cycles between kits (Aikin 1990b).

Once programmed, the drumKat differs from orchestral percussion in how the player cognizes it. Time-variant behavior leads to questions like what the next sound will be from a zone cycling through sounds. Force-variant behavior heightens awareness of impact force, when crossing a threshold can suddenly produce a change of timbre. Finally, with physical percussion the player has obvious visual and spatial cues when walking or turning from one collection of instruments to another, while with the drumKat a change of kit is confirmed only on a tiny digital display. Such changes, if frequent, demand considerably more conscious attention because of the lack of sensory feedback supporting them.

Even more versatile than the Zendrum and drumKat is Don Buchla's Thunder, intended for playing with bare hands instead of mallets. Many zones (25 of 36 in all) are sensitive to strike velocity and after-strike pressure. Ten small zones across the top are used for menus altering the behavior of the instrument, *e.g.*, remapping which gestures get sent to which MIDI sounds. A narrow strip near the top of the surface functions as a ribbon controller. The remaining large zones sense impact position and can be divided into subregions.

Like most of Buchla's designs, Thunder is extensively programmable. Subprograms consisting of transpositions, delays, repeats, amplitude envelopes, and parameter crossfades can be activated for particular MIDI channels; each subprogram can be enabled or disabled by any zone. Zones can initiate, terminate, or toggle arbitrary notes. Nine of the zones have binary state indicated by an LED; this state can be set, reset, or toggled from any zone. Impact speed, impact position, and after-strike pressure can be assigned to arbitrary MIDI commands (Aikin 1990a). For specialized use, as much time may be spent by the composer/instrument builder programming and evaluating designs on the Thunder as the performer spends rehearsing. And certainly designs could be made which exceed what can be kept track of during performance: the Thunder is a prime example of an instrument beyond the performer's full comprehension during performance (and, like Babbitt points out, one which can control sounds beyond a listener's discrimination).

Some percussion controllers extract more information from their pads than mere impact force. If the pad resonates at all, the vibration measured by the transducer on the pad is susceptible to analyses beyond simple peak detection. Korg's Wavedrum implements this, using the measured vibration as an excitation signal for its various synthesizers (Harenberg 2000). The "bonk" algorithm recognizes individual percussion attacks in a continuous sound stream by measuring its frequency content instead of merely following its amplitude envelope. This method is particularly robust when a resonating pad reports false secondary attacks, or a when rapid sequence of attacks produces spectral change but no amplitude change (Puckette, Apel, and Zicarelli 1998).

2.2.1.5 Celletto

Chris Chafe's electric cello, called the celletto, has a bodyless frame supporting the fingerboard and bridge, with contact points for knees, chest and arms. Piezoelectric pickups on each string capture the acoustic signal. An accelerometer and strain gauges mounted on the bow measure bow speed and force. The celletto has also used a Buchla Lightning infrared motion tracker to track the position of the frog of the bow in the vertical plane (this is practical because the cellist is stationary; bow and instrument are always visible from a fixed point in front of the performer).

Pitch and amplitude are tracked and converted to MIDI commands with the Period Predictor Pitch Tracker (PPPT) introduced in (Cook 1991). When tracking a trumpet, presumably in its middle register, its average latency is reported as 30 msec (Cook 1993). This number is a lower bound on the celletto's latency because the range of the celletto is lower than that of the trumpet, and also because the tracker in (Cook 1993) helps predict the pitch by sensing the position of the trumpet's valves. Such a hybrid approach is lacking in the celletto. So Chafe uses the pitch tracking more for score following (triggering events, transposing other material) than for controlling the pitch and amplitude of a synthesizer; he does not attempt to work around the latency inherent in tracking the low pitches possible with the cello (Chafe 2000).

As bow position, speed and force are explicitly tracked, these parameters naturally command extra attention on the celletto when compared to the cello. This is compensated for by the lesser amount of attention directed towards good tone—the signal coming from the bridge often undergoes heavy processing which masks what cello players consider subtle aspects of tone quality. As with the WX-7 some aspects of the base instrument are extended; other aspects are then necessarily restricted, at least by the player's finite attention if not by explicit design decisions.

2.2.1.6 Violin controllers

Over a dozen companies sell electric violins, Zeta being the most widely known. Performers often prefer electric to acoustic violins just because they resist feedback when amplified; otherwise they play them conventionally. Some electric violins have up to three additional lower strings (in fifths or in standard guitar tuning) or fretted fingerboards to allow transfer of technique from other fretted instruments. Manufacturers like Steinberger, Jordan, and Jensen exploit its new design constraints by eliminating the body, moving the tuning pegs to the chin, using lighter and stronger materials, and altering the geometry of shoulder and chin rests.

An electric violin can be called a synthetic instrument when its signal is converted into data, making it possible to separate input gesture from the output sound and thereby alter the mapping connecting them. It allows idiomatic violin bowing and fingering gestures to be mapped to arbitrary MIDI-synthesized sounds. MIDI information sent from these controllers usually consists of note-on, note-off, pitchbend, and variation of amplitude. Some instruments can drive a separate MIDI channel from each string; less expensive models have only one pickup for all the strings, which requires polyphonic pitch tracking when double-stopping or even when letting open strings ring unintentionally.

It is possible to track additional timbral parameters such as spectral content (overall brightness, energy per critical band) and rate, depth, and evenness of vibrato. Spectral parameters would function as a group, as would vibrato-related parameters. Vibrato has an inherent latency of tracking so it would best be used to control aspects of synthesis where slow response is acceptable. Vibrato is so automatic for most players, though, that it makes sense to not merely ignore it or leave its effect implicit.

The downward force exerted by the bow on the strings can be measured by mounting sensors on it, for instance force-sensing resistors on the R-Bow of Dan Trueman (1998) and strain gauges on the celletto. The R-Bow also uses accelerometers to estimate the linear position of the bow relative to the (presumed stationary) bridge, and to estimate which string is being played by measuring the angle the bow is held at.

An extreme example of a violin controller is the BoSSA synthetic instrument, a “deconstructed violin reconstructed” (Trueman and Cook 1999). The two pressure sensors and two accelerometers on its R-Bow provide four dimensions of control; its fingerboard simulates only a single string, but six more dimensions of control come from measuring its angle (mapped to pitch bend, analogous to bending a guitar string) and the finger pressure on four pads mounted on it. The bow excites four “sponges” rather than four strings, which effectively measure bow speed and distinguish up-bow from down-bow. Trueman finds that about nine simultaneous dimensions of control are available in performance. A simple mapping from these controls to sound is given in his composition *Lobster Quadrille*. The four strings correspond to four phrases from Lewis Carroll’s poem; bow direction determines direction of sample playback and fingerboard technique determines pitch. The speech samples are sent through effects processors: rotating the fingerboard sets the amount of processing and bow pressure applies a vibrato to the processing. Finally, the R-Bow accelerometers drive a model of wind chimes, so the player can shake the chimes directly.

2.2.1.7 The Hypercello and Hyperviolin

Researchers at MIT's Media Lab, under the direction of Tod Machover and with the technical guidance of Joseph Paradiso, have designed many instruments. The Hypercello is the first of Machover's "hyperinstruments," dating from 1991. A radio signal broadcast from an antenna above the bridge is received by a resistive strip running the length of the bow. Induced currents flow from each end of this strip; their difference and sum yield the transverse position of the bow and its distance from the bridge respectively. A deformable capacitor mounted under the right index finger indirectly measures the downward force exerted by the bow on the strings (Paradiso 1998, 1997, 1999c). An exoskeletal device measures the angle of the right wrist, yielding more data about bowing gestures. Left-hand finger positions are directly measured with resistive strips on the fingerboard. Finally, the audio signals from each string are analyzed with software.

Excluding the audio signals, we estimate 90 bits of data per frame are sent from the Hypercello. The sending microcontroller and receiving host computer, a Macintosh IIfx, could maintain a serial data input rate of 9600 bits per second (Machover 1997); so the latency for measurements relating to fingers, wrist and bow is at least 10 msec. As the audio signal is apparently not used for pitch extraction, overall latency can be kept quite low, probably under 20 msec. Standard MIDI 7-bit resolution suffices for these spatial measurements.

The Hyperviolin (1997) uses a wireless bow tracker. Two radio transmitters and a resistive strip on the bow work with a receiver atop the bridge, like the bow tracking of the Hypercello. A force-sensitive resistor under the right index finger sends its measurements to the host computer via a third radio transmitter.

2.2.1.8 A muted trombone controller

Mark Bromwich's Metabone is a trombone practically silenced with a custom foam-filled mute, similar to the practice mutes available for most brass instruments. The end of the trombone's slide interrupts one of two ultrasonic beams to generate two distance values. Short-range capacitive sensors, inductive sensors, and infrared sensors are also reportedly used, though Bromwich (1997) does not state precisely what is sensed, bell, slide, or performer's body; nor does this dissertation specify mappings from sensors to sound.

2.2.1.9 The HIRN controller with WhirlWind physical-model synthesizer

The WhirlWind synthesis algorithm is a physical model combining reed, lip, and jet excitation models (clarinet, trumpet and flute). It is controlled by MIDI commands from a physical controller called HIRN, shown in figure 13 (Cook 1992). HIRN provides four spatial controls. This is less general and less flexible than the twelve degrees of freedom afforded by free motion of both hands, but is accomplished with far simpler technology than full motion tracking. (The perceptible friction and limited range of motion are other advantages of this construction. Instruments such as the Theremin (see below) which sense the position of the performer without direct contact are difficult to play because of their lack of tactile feedback.) "The right hand can be slid linearly along the axis of the instrument [9], as well as radially rotated [8]. The left hand can be rotated radially [6]. Finally, the head of the instrument can be rotated [5]" (Cook 1992, 275). The mouthpiece measures breath pressure, bite tension, lip tension, and pitch/amplitude of the performer's humming, five more scalar controls. Finally, the fingers operate eight woodwind-style keys. MIDI foot controllers are optional.

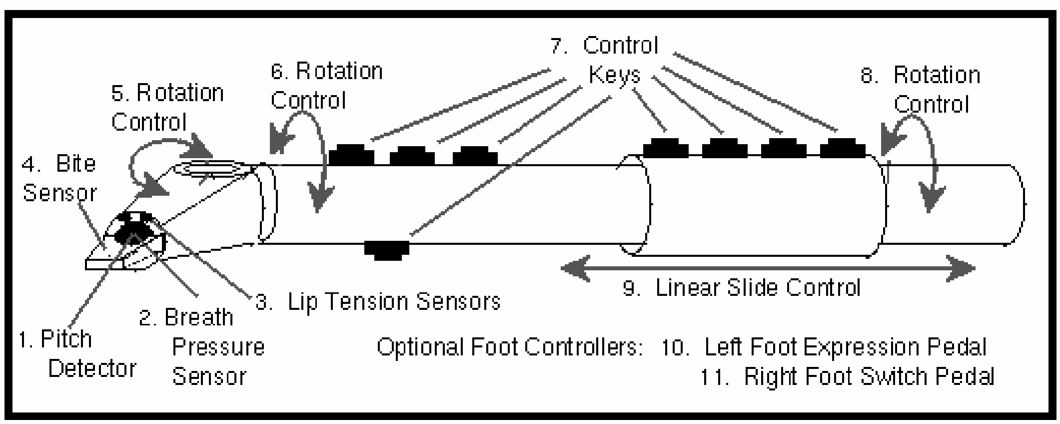


Figure 13. Perry Cook's HIRN controller. Reproduced with permission from (Cook 1992).

Cook suggests the following mapping from HIRN controls to WhirlWind parameters, mostly one-to-one. Breath pressure, lip tension, and bite pressure map directly to their corresponding elements in WhirlWind. Rotation of the head joint drives the synthesizer's balance between reed and jet excitation (clarinet and flute). Rotation of the left hand drives flute embouchure. Right hand slide drives pitch (one octave) by adjusting the length of a delay line, with two fingered keys to select one of four octaves (somewhat like a trombone). Right hand rotation drives the amplitude of noise injected into the synthesizer. Left foot pedal drives overall output volume. Right foot "damper pedal" acts to sustain the currently measured breath pressure.

2.2.2 Distance sensors

2.2.2.1 Theremin

The Theremin produces an approximately vocal spectrum and is most easily played in the pitch range of a soprano singer. Pitch is driven by the distance of the right hand from a vertically oriented antenna, 2 kHz at closest approach, approaching 0 Hz when the right hand is held far away. Clara Rockmore, arguably its best performer, adapted her violin skill to the instrument by holding her right hand in different "positions" and playing various pitches by moving her fingers while holding the hand steady (Rhea 1989). Resolution of pitch is essentially continuous, as the instrument uses analog electronics; certainly it exceeds the limits of human pitch distinction. Latency of pitch change is also below the threshold of human perception.

Amplitude is the only other controllable parameter. The left hand approaches a second antenna (actually, it passes through a loop) to reduce amplitude from a fixed maximum to silence. There is a slight lag in diminuendo as it takes some time for the amplifying vacuum tube to cool down (this is the reason for the vent at the top of the instrument), from 100 msec for beginner models to near zero for advanced ones. This is analogous to an instrument with a physically vibrating body, whose energy also takes time to dissipate when its vibration is damped.

2.2.2.2 Radio distance sensors

The theremin measures the position of the human body directly. Some of the instruments below do so too; others require the performer to hold objects whose positions are measured.

The Mathews/Boie Radio Drum, also called the Radio Baton, tracks the three-dimensional position of the heads of two mallets, attached by wires to the main unit. Schloss (1990) explains that each mallet head contains a coil of wire which transmits a low-frequency radio signal to the drum surface, which is divided into a rectangle surrounded by four thin strips (total size about 26 by 36 cm). Position measurement is derived from the first moment of capacitance between the coil and the surfaces. The positional accuracy of the Radio Drum is about 5 mm when the mallet head is near the surface; its latency is typically 5 to 10 msec. Boie and Schloss (1989) note that measurement of height is monotonic but decidedly nonlinear. It is possible up to a height of about 15 cm; with increasing height, positional accuracy decreases. The radio sensing technology could easily work at sub-millisecond rates; the bottleneck of speed is actually communications to the host computer. They distinguish momentary and continuous control modes. Firstly, triggers can be set based on height: a trigger is fired when the height falls below some threshold, and further fires are disabled until the height exceeds a different, higher threshold (this hysteresis prevents double-triggering). Secondly, they describe continuous control of timbre by mapping the coordinates of a mallet head onto spectral shape, envelope, and vibrato. They surmise (and this is over a decade ago!): "We suspect that the usual 'synthesizer timbre' is caused by limitations in traditional control devices—keyboards, footpedals and knobs—and not by inherent limitations in sound synthesis hardware" (Boie and Schloss 1989, 45).

Several instruments which perform remote sensing with radio fields and electric fields have been built at MIT's Media Lab. We describe three of them here.

The Sensor Chair uses electric field sensing in "transmit mode" to measure the position of the hands and feet of a seated performer. The performer's body is attached to an electrode and transmits an electric field; sensors pick up this field strongly or weakly as the body moves closer or farther from them. In one mode, the performer can trigger notes by moving a hand forward, then changing their timbre by moving the hand laterally. In another mode, moving hands and feet trigger percussive sounds, spatially disposed like a conventional trap set. A lamp is mounted beside each electric field sensor to provide visual feedback of the instrument's behavior, for both audience and performer (Paradiso 1998, 1997).

The Sensor Frame uses another mode of electric field sensing, shunt mode, which requires no electrodes. The human body is placed between transmitters and receivers, shunting the field to ground rather than letting it pass. A typical configuration is a 75 cm square frame mounted vertically like a window, with two transmitters at opposite corners and many receivers along the edges. The position and size of a hand (or other body part) in the frame can be inferred from this arrangement; with simplifying assumptions, more than one hand can be tracked. Horizontal and vertical position of the shunting body part typically control the pitch and timbre of a MIDI instrument.

The Magic Carpet consists of a 10 cm grid of piezoelectric wires running underneath a carpet 4 m square, combined with a pair of microwave Doppler radar motion sensors. The grid of wires measures the pressure and position of the performer's feet, while the motion sensors measure the speed and direction of the torso in the horizontal plane (Paradiso 1999b). From seeing performances of it, we estimate the latency and frame duration of both sensors to be under 50 msec. As both sensors commonly trigger sounds and sequences from MIDI synthesizers, resolution is probably 7 bits.

2.2.2.3 LED distance sensors

The Twin Towers controller senses the position of two hands immediately above a platform, to a height of about 50 cm (Tarabella 1997). Each hand reflects light from an upward-pointing infrared LED back down to four infrared detectors, mounted in a square around the transmitting LED. So at its most basic level it is like a bank of eight sliders. Software typically converts the raw height data into discrete gestures (angle of hand, overall height, off-center position, speed of motion, *etc*.) particular to individual compositions. With more detectors this could conceivably measure all six degrees of freedom of each hand, over a small physical space. Latency is about 50 msec, though it could theoretically be under 1 msec.

The most recent version of Don Buchla's Lightning controller uses infrared LEDs to track the two-dimensional position of two wands in a vertical plane, reporting these as MIDI control messages. Transitions of two momentary switches on each wand are also reported. It also recognizes certain gestures and reports those as MIDI note-on messages. The vertical plane can be divided into zones, each zone with its own mapping from wand position to MIDI commands.

The MIT Media Lab's Digital Baton is shaped like a largish conductor's baton. The horizontal/vertical position of its tip is tracked with infrared LEDs for precise pointing. A triaxial accelerometer quickly detects directional beats and large gestures. Five force-sensitive resistors in the baton's skin measure finger and palm pressure. As 8-bit resolution is appropriate for these data, and as they are sent by a microcontroller over a serial cable to the host computer, a 9600 baud connection would yield a frame duration near 10 msec. Overall latency we estimate to be 20 msec. The Digital Baton is effective as a modulator of fixed musical sequences. Beat patterns can add accents; finger pressure can adjust speed; beats pointing forward can effect higher-level structural changes to the playback of the sequence. It can also be used as a more traditional instrument, where (for instance) vertical position drives pitch and horizontal position drives some timbral parameter (Paradiso 1999a).

2.2.2.4 Video-capture devices

A sub-family of distance sensors uses consumer video equipment and image processing software as the sensing element. Latency is too large for fine control of musical gesture, being at least 33 msec because of the fixed video broadcast standard of 30 frames per second. The small number of scan lines in a video broadcast signal affords limited spatial resolution, too. The advantages of these devices are low cost, good scalability for tracking multiple objects as image processing software improves, and impressive stage presence. If only a few objects are tracked, though, actual acoustic control is generally worse than with other technologies.

Dozens of objects could in principle be tracked in three-dimensional position and orientation, particularly if several synchronized cameras provided different viewpoints, with no more equipment needed than for tracking one object. Software to do this still requires significant expertise in pattern recognition to use effectively, though. A simpler two-dimensional high-resolution video controller has been delightfully demonstrated at recent SIGGRAPH conferences, where about a thousand persons in a darkened auditorium use two-sided red/green retroreflective wands to "vote" one way or another from moment to moment. This lets the whole audience collaboratively play videogames, fly airplanes, and perform other tasks taking two scalar inputs, measured by counting votes in the left and right halves of the auditorium.[80]

Tarabella (1997) describes four video-capture devices. (i) The visually tracked component of the Aerial Painting Hands is two gloves with colored dots on the palm and back. Each glove's attitude (hand open/closed) and position in the vertical plane (like the Buchla Lightning) is tracked. (The colored dots cause corresponding colored lines to be drawn on a projection screen, but this does not affect the sound.) Closing and opening the right hand triggers a corresponding command for the sound synthesizer, with parameters the two-dimensional position of the hand at that time. (ii) The UV Stick tracks the two-dimensional position and orientation in a vertical plane of a 50×3 cm wand which fluoresces under ultraviolet illumination. (iii) The Imaginary Piano tracks the fingers of a pianist without a piano. As a finger crosses below the imaginary line of a keyboard, a MIDI note-on command is sent for the pitch corresponding to that position, with velocity corresponding to the speed of the finger. (iv) The Light Baton System is a handheld wand with a light at its tip. Its gesture recognition software is tailored to recognizing events happening with particular parameters (size or speed of a part of the gesture).

We can generalize from these four examples: video capture can be used to continuously control several devices; detection of momentary events can be used like a percussion attack with some set of continuous parameters derived from details of the event (its position, speed, and color).

Video camera systems are popular for dance interfaces too, notably the Very Nervous System (Rokeby 1995; Winkler 1997).

[80] Shouting from the audience (and, since 1998, cheating by indicating strategy on the large projection screen with laser pointers) is an integral part of such a collaboration.

2.2.2.5 Measuring joint angles

The DIEM Digital Dance System measures up to 14 joint angles on a dancer or musician by means of bending sensors, with latency of about 30 msec. The data is presented for arbitrary interpretation by MIDI hardware (Siegel 1998; Siegel and Jacobsen 1999). Ivar Frounberg's composition *Drømmespor* uses DIEM to infer the vibrato and bow movement of a violinist from motion of elbow and wrist. Myoelectric sensing, two skin sensors placed on opposing flexor and extensor muscles, can be used to infer joint motion instead of measuring it directly. This is particularly popular for dancers, since the sensors do not inhibit movement. Gillett, Smith, and Pritchard (1985) describe such a system which transmits the sensed values via radio to the music system. They also suggest that dancers' control of the music be limited to improvised arrangement of precomposed fragments. Trying to play a full-blown musical instrument while dancing would likely impose unsatisfiable constraints on a single performer. Wilson-Bokowiec and Bromwich (1999) also support this idea; their Bodycoder suit sends 14 channels of MIDI data to control sequences of visual images and playback of audio samples.[81] The audio samples can be modified with pitch changes, filtering and enveloping. But the recognizability and nonneutrality of most sampled sounds stands in marked contrast to the less fixed emotional associations of a family of sounds produced without samples (such as that produced by orchestral instruments).

2.2.3 Other controllers

2.2.3.1 Continuous keyboards

Haken et al. have worked for some years designing polyphonic controllers with independent continuous control of pitch, loudness and timbre. Their most recent instrument, the Continuum, tracks the position of up to ten fingers as well as the downward force of each finger on the surface. Lateral position maps to frequency, downward force to amplitude, and forwards/backwards position to an interpolated cross-synthesis between trombone and cello. One constraint on polyphony is that multiple notes of the same pitch (the same lateral position) cannot be played; they will be collapsed into a single note. Finger positions are measured to an accuracy of 1 mm and are recomputed every 4 msec, to distinguish fast glissandos from new note onsets (Haken, Tellman, and Wolfe 1998).

The Continuum cross-synthesizes tones from a collection of timbral analyses of trombone and cello tones, at various levels of pitch and loudness. These analyses are arranged in a 3-dimensional lattice much in the spirit of (Bowler et al. 1990). The lattice has only two points in the timbral dimension (trombone and cello), three in loudness, and about a dozen irregularly spaced points in pitch. The surface position and pressure of a finger are considered to be a 3-dimensional point in the space bounded by the lattice, and particular synthesis parameters for that point are found by trilinearly interpolating between the analyses corresponding to the 8 vertices of the particular cubical cell which surrounds that point (a generalization of bilinear interpolation, covered in the next chapter).

Single melodies are fairly easy to play accurately when engaging a secondary control which rounds frequency to the nearest tempered semitone. But playing multiple notes on the Continuum is quite difficult. Reliably playing a single ten-note chord, never mind a three-part fugue, requires hours of rehearsal even for an accomplished pianist.

The earlier Rolky keyboard tracks the position of up to five fingers on a plexiglass sheet with a video camera (Johnstone 1985), so its latency is considerably higher than that of the Continuum.

Snell (1983) proposes a keyboard intermediate between discrete and continuous: black keys which slope down at the back, leading to a pressure-sensitive flat plate which allows continuous pitch gestures. The 1948 version of Hugh Le Caine's Electronic Sackbut implements a simpler design above its keyboard: a pressure-sensitive strip, actually keys one seventh as wide, controls pitch nearly continuously (Young 1989, 173). As another compromise for accurate intonation on continuous keyboards, we can imagine a dead zone around the precise spatial location of each tempered semitone. A more advanced design would constrain notes to begin exactly on pitch but begin to bend away once the finger had moved a distance greater than some threshold. (All of this equally applies to tunings other than 12-tone equal temperament.)

2.2.3.2 Biological sensors

Since the 1960's devices have been built to measure subtle and normally involuntary aspects of the human body. The BioMuse (Knapp 1990) is a recent model, measuring 8 channels of electroencephalographic and electromyographic data at a sampling rate of 4 kHz. It emits serial data at 19.2 kbaud (about 2000 bits per second) or MIDI data. We generally exclude such devices from our discussion of synthetic instruments because the role of performer is so different with them. Tracking the amplitude of alpha, beta, and theta brain waves, galvanic skin response, body temperature, rate of breathing or heart beats has been successfully used in compositions, notably by David Rosenboom (technology, aesthetics and compositions described in (Rosenboom 1976; Rosenboom 1990)). But even when performers have been

[81] The Bodycoder suit is built around force-sensitive resistors (FSR's), the most durable and reliable sensors they found which dancers could wear.

extensively trained to consciously control these aspects, their gestures are slower, have less temporal precision, and have more limited range than conventional muscular actions. Their interest lies in letting the performer relax control and unconsciously react to the environment (including the sounds they are causing). This in part rejects our initial premise that there be at least one performer: there may be a partial performer or no performer, in the sense of one who acts voluntarily and intentionally.

2.3 Summary of analyses

Generalizing grossly, the controls of an orchestral or synthetic instrument affect sound primarily in its pitch, secondarily in its amplitude, lastly in its timbre. Amplitude is sometimes constrained by choice of pitch; in many cases timbre is almost entirely determined by pitch and amplitude. This state of affairs comes in large part from Western common-practice harmonic tradition, where compositional structure derives far more from harmony and melody than from dynamic level and tone color.[82] So even with unpitched percussion we often find a family of objects (wood blocks, cymbals) sorted by size, *i.e.*, pitch. Timbre is often the most difficult thing to control, notably for beginners on orchestral instruments. More commonly than with orchestral instruments, synthetic instruments have separate controls for timbre (a weaker cross-coupling of timbral parameters with other parameters); but these controls are rarely as powerful as pitch controls, and divert the performer's attention (if not an entire limb). And when these timbral controls are ignored, the static timbre is tiresome. Again generalizing, those instruments which offer the most timbral control—percussion and voice—also offer the smallest pitch range. Controls and attention otherwise directed to precise control of pitch over a large range are here made available for timbre.

2.3.1 Acoustic parameters

In this summary we divide acoustic parameters into the three broad areas of pitch, amplitude and timbre. Pitch is usually driven by controls which are proportional with respect to musculature. Controls of higher order, such as integral controls, lack agility; controls of lower order such as derivative controls make it too hard to maintain a steady pitch, a requirement for harmony.[83] Hand and arm work together (strings, claviers), or both hands work together (woodwinds). If switches control pitch, they are almost always momentary rather than latched; but electronic instruments sometimes use latched switches for

[82] Rhythm is largely independent of these factors.

[83] The eviolin configured with bow speed driving pitch and fingered pitch driving amplitude is unplayable: wide uncontrolled variation in pitch completely distracts the ear from hearing subtle controlled variation of amplitude. Only by reducing the pitch range to a minor third can the control of amplitude be made audible in this configuration.

controls that maintain state for extended durations like octave switches. The primacy of pitch is reflected in the often multiple controls which drive it, particularly in sustaining instruments like woodwinds and strings.

We observe that amplitude is usually driven by a derivative control (speed, not position, of bow, diaphragm, or mallet). We can explain this in terms of both control theory and classical physics. Controls with orders greater than −1 (proportional, integral, and beyond) are less agile, less able to quickly change value. Granted, higher-order controls can more easily maintain a constant value, but this is not as important for amplitude as it is for pitch. On the other hand, controls with order less than −1 are so agile that they are generally too unstable to be played. This leaves −1 as the happy medium.

From physics we know that the energy of a vibrating object (or wave) is proportional to the square of the vibration's amplitude. Since this energy is purely kinetic at the vibration's point of rest, we can apply the equation of kinetic energy to show that energy is proportional to the square of velocity. (Energy at this point is correlated with only speed, not position or acceleration.) Combining these two facts about energy ($A^2 \propto E$, $E \propto mv^2$), we conclude that the amplitude of a sound should vary linearly with speed ($A \propto v$) in order to agree with the performer's everyday physical experience; amplitude varying with acceleration or position is counterintuitive. This is confirmed by experiments driving eviolin amplitude from bow position, velocity, or acceleration: effective controls of amplitude have order −1 (with respect to position).

Rapid subtle control of timbre was not common in compositions when most orchestral instruments were being developed, so timbre is often relegated to secondary controls in these instruments. Brass mutes, *sul ponticello*, *col legno*, and string mutes exemplify such controls. (Since then, composers and performers have taken up the challenge of using these controls as primary, as with Vinko Globokar's virtuoso trombone muting in his *Echanges* (1973) and other compositions.) Timbre is often controlled in only one dimension at a time, extending from *modo ordinario* towards some unplayable limit. Orchestral examples of this are *sul ponticello*, *cuivré*, and breathier tone. Commercial synthesizers reflect this trend too: many have only a single "mod wheel."

During the attack portion of a sound, high-frequency energy is often proportional to the tension of muscles immediately before their release allows a sound to begin. The sudden release of energy is a transfer from muscles to sound-generator (at least for non-electronic instruments). Typically a greater amount of energy input into a system causes disproportionately more energy to be transferred to its higher-frequency modes of oscillation. This is so because there is simply more room for different modes of oscillation in the higher part of the spectrum. Then as energy dissipates after the initial transfer from muscles to system, it dissipates more quickly from these higher-frequency modes. We conclude that if a gesture begins a sound, greater energy in this gesture should boost the high frequency content of the sound. This is again to better agree with the performer's everyday physical experience.

2.3.2 Gestural inputs

The controls of an instrument often use the fingertips, lips and tongue. These parts of the body are exactly those with finest motor control and greatest sensitivity, both proprioceptive and external.

Lighter parts of the body have less inertia and thus can move more quickly and accurately. In some cases the hands hold tools to simulate body parts with even lighter weight (and extra length), as with mallet percussion instruments. Fast high-accuracy musical gestures (typically pitch) are assigned to lightweight parts like fingers, upper arms, mallets, and mouth; slower gestures can be assigned to large arm motions, feet, legs and torso. Especially for players who stand, posture and balance must be respected when assigning gestures to the larger body parts.

Mapping a position to a proportional control is simple, but playing for a long time in a nonstandard position or orientation can be tiresome and inefficient. This can be avoided in several ways. Most easily, the control can be changed from proportional to integral, possibly with a dead zone surrounding the zero midpoint. Then nonzero positions need only be held for short durations, like the steering wheel of a car stays turned for only a short time. If this integral control turns out to be too agile, it can be left as a proportional control modified in the manner of a thermometer which displays its maximum past reading. The most extreme position so far, instead of the current position, drives the control's value. So again performers need not maintain an extreme pose (though they can if they want to for dramatic effect). A separate gesture is then required to return the control to its original value. This technique can work in one direction (from one extreme of a parameter towards another) or two (outwards in both directions from a central value).

The findings of Vertegaal, Ungvary, and Kieslinger (1996) can be summarized concisely with this terminology. Effective isometric (scalar) controls are integral with respect to applied muscular force; effective isotonic controls are proportional with respect to position. Isometric controls rely particularly on tactile feedback, how hard the control is pushed, rather than visual feedback. A discrete control works best when its different states correspond to different spatial positions, not different speeds or forces; visual feedback can play a greater role here as well.

Some percussion instruments offer more control of timbre than sustaining instruments. This seems paradoxical, because a sustaining instrument can change its timbre from moment to moment while a percussion instrument generally specifies its timbre only during the brief instant of impact. The explanation lies in what leads up to impact. Most struck objects move a little and offer varying resistance to the mallet—drum heads, chimes, gongs, even piano keys. The tension in various muscles immediately before and during the noninstantaneous moment of impact determines the interaction (*i.e.*, energy transfer at various frequencies) between mallet and struck object. The grip, wrist, elbow, shoulder all can vary in stiffness from attack to attack. These various tensions are set up before the sound itself begins; more attention can be given to these tensions in this "set-up mode" than would be possible if these tensions continuously controlled the sound. Also, effective mallet length can vary from attack to attack; with multiple mallets in one hand, weight and hardness of the mallet head can also vary, albeit discretely.[84]

In the extreme case, drums struck with the hand like the tabla and dumbek produce an enormous range of timbre. Tabla technique begins with a range of attacks from flicks and live strokes through dead strokes and finally thuds. Feldman (1983) accessibly introduces the mechanics and Indian terminology for the tabla. These attacks vary in the total inertia of the effective "mallet" (fingertip, flat of the index finger, multiple fingers together, knuckles, flat of the hand, palm) and in the amount of damping after the attack, by the mallet itself or by a separate entity like another part of the hand. Damping by pressing on the drum head can raise the pitch for part or all of the stroke; this is common for the lower-pitched left-hand drum where the palm always rests on the drum head. A separate timbral dimension independent of liveness of stroke and loudness is given by the location of impact. The thick center of the drum head produces a strongly pitched sound; its surrounding ring gives a resonant sound, yielding to a sharper attack on the outer part of the drum head; the rim itself produces a much drier tap. All these timbres are commonly combined by striking several regions at once with various hand shapes. Finally, these strokes are combined in time: a variety of grace notes are performed with multiple fingers, and left- and right-hand strokes are combined in ways similar to Western marching-band rudiments into composite strokes.

Such a variety of technique has bred a special terminology. Each stroke is given a monosyllabic name (only loosely onomatopoetic), and these syllables combine into words. This artificial language efficiently encodes a large amount of information and is indeed necessary for teaching. (Foreign students are strongly encouraged to learn Indian pronunciation.) Cognitively, we suspect that for a non-tabla-playing musician (say, a pianist) to play a tabla synthesizer well, the instinctive "language" of the pianist—chord shapes, chord voicings, varying tessitura—should be mapped in structure-preserving fashion onto this language of tabla.

---

[84] A bow striking a string has similar complex muscular control. We can draw analogies between various bow techniques and various percussion attacks; indeed, nonlegato string playing has a significant unpitched noise component at the start of a note.

At another extreme, some pianists joke about it not mattering if a piano key is depressed by Paderewski's finger or by the tip of Paderewski's umbrella. But this exception sort of proves the rule. Some gestures (chord voicings, phrasings in fast passagework) which I cannot perform on excellent electronic pianos, I can reliably perform on a conventional grand. This may be because the piano action responds to the *varying* force applied by my finger during the key's travel; the speed of the key at the end of its travel determines how fast the hammer flies towards the strings. Commercial keyboard controllers measure only the duration of the key's travel: the average speed, not the final speed. The information lost by this simplification is the subtle variation of muscular tension from finger to finger in a chord or run.

Multiple modes of excitation are possible with electronic instruments. Several guitar controllers use clavier-type keys as well as plucked strings to initiate a note, offering two families of gestures instead of one. The HIRN controller goes farther, offering three brass and woodwind modes of excitation simultaneously and with continuous interpolation. Some percussion controllers similarly combine the measuring of attack strength with continuous measurement of after-attack pressure (or even the spectral content of the pressure transducer's raw output) as a multiple-mode input to the synthesizer.

When several muscles drive a control, particularly in the case of derivative control of amplitude, the performer can make the sound seem infinitely sustained by alternating between two sets of muscles. Examples of this are circular breathing (alternating between cheeks and diaphragm) and bowing a single vibraphone note with two bows. Less magically, a nonsustaining instrument can emulate a sustaining instrument if it can be excited rapidly and uniformly, like shaking a rattle, playing a drum roll, or playing a mandolin *tremolando*.

Additional structure can be imposed on a homogeneous control; here are four increasingly elaborate examples. (i) A pad on the Zendrum can classify how hard it is hit into one of several ranges. (ii) MIDI keyboard controllers are commonly divided or "split" into regions with different timbres. (iii) The Buchla Thunder's surface is divided into zones with quite different functions. (iv) Finally, computer displays are often subdivided into menus, dialog boxes and buttons all accessed with a single mouse.

On some controls it is easy to play an alternating pattern: a trill, or an alternation between a fixed value and a varying melody as in the previously cited final movement of Beethoven's Violin Concerto. Some woodwind trills and most string trills (and trills on the *ondes martenot* and early monophonic synthesizers) fall into this category. The fixed value can be thought of as a movable detent.

2.3.3 Controls

The performer can more easily play simultaneously several members of an array of controls when the resolution or number of degrees of freedom of each individual control is reduced. We see two examples of this compromise between high-level and low-level control in comparing fretted and unfretted fingerboards, or conventional and multiply touch-sensitive claviers. This compromise, reducing the strength of individual controls in an array, has traditionally been fixed by the instrument builder. But software could adjust the strength from moment to moment in a performance, either by measuring how many controls are active or by explicit command from the performer through a secondary control. As a simple example of this, one configuration of the eviolin uses transverse position on stage to determine how much to round off pitch to the nearest semitone. In fast passagework the eviolinist stands where it completely eliminates poor intonation, while in more lyrical sections he can stand where he is free to slide between notes and use much vibrato. But it would be even nicer if the instrument (accurately!) classified how lyrical the playing style was from moment to moment, and reacted accordingly.

Another compromise for arrays of similar controls is simply spatial. The smaller each control is, the more controls lie within reach of the hand or arm. The best result lies midway, not at an extreme. A huge array of tiny controls offers large reach but this is more than offset by the difficulty of manipulating or even selecting the individual controls (a violin with a dozen strings on a standard bridge, or a piano with half-width keys).[85] Conversely, a very small set of easily manipulable controls becomes indistinguishable from only one control. The "impedance" between controls and performer must be matched.

A momentary switch can augment the behavior of another control by acting as a latch, sustaining the value which the control had when the switch was depressed. This is particularly useful for derivative controls, which cannot maintain an unchanging value for an extended duration. Because of its sustaining nature and the familiarity of many performers with the damper pedal of the piano, such a switch is appropriately assigned to the right foot.

MIDI pedal keyboards do not extend performance possibilities beyond what is already found in organ and pedal piano. They seem to be intended for non-keyboardists to play simple bass lines while playing another instrument; in this realm, they suggest the possibility of performing on several instruments at once. Of course several players could play these instruments, too. And several players can play one instrument. Several can even play on several, nondecomposably: the Willem Breuker Kollektief has had

[85] Rescaling the pitches played by a violin controller so fingered quarter-tones sound as semitones is a continuous analogy: increased range requiring increased accuracy.

one performer blowing two saxophones, fingering one, the other saxophone fingered by a second performer standing behind the first.

Because the ear distinguishes very fine differences of pitch over a wide range, a single continuous control often does not suffice to drive pitch. The full continuum (or as much of it as is accessible to a given instrument) is often divided into regions ("modules" according to Boulez, though he treats problems of composition, not instrument design). This modularity is reflected in the structure of a conventional clavier, repeating a pattern every octave. One control works within a region, the other selects one region from the available ones. Proficiency on an instrument often involves using both these controls fluidly, making it sound like only one control: scales and arpeggios across strings or across woodwind breaks, or almost any angular melody on a brass instrument. Exercises for crossing strings and crossing the break are common. The WX-7 multi-octave key also demonstrates this design.

The additional trill keys of the WX-7 make certain trills much easier (or possible at all). This shows how redundancy can be valuable in an interface for a difficult control task like pitch. This pitch control is here at a very small scale, not over the whole pitch gamut; but it relates to dividing the gamut into regions in that the otherwise difficult trills often cross the boundaries between these regions.

2.3.4 Notation

A large part of the notation of pitched orchestral instruments is common to them all: the five-line staff, note heads (various kinds indicating various durations), clefs, accidentals, dynamic markings, textual instructions for extraordinary modes of playing, phrase marks, Italian adjectives. We consider the implications of all these for the composer's mental model of an instrument and its performer.

The staff has horizontal and vertical extent, and horizontal and vertical structure. Most fundamentally the staff marks the passing of time; the composer thus considers the instrument as acting in time, and the instructions he gives the performer as the placing of events in time. The second thing the staff delineates, its vertical dimension, is pitch. Thus, before all else, the instrument is conceptualized by the composer as a device for tracing a pitch contour through time.[86] In the vertical structure of the staff we find a diatonic structure, the letter-names of the notes. The farther a composition reaches from classical tonality, the poorer this notation is; but even a mathematical dodecaphonist like Babbitt is forced to abide by it in concession to his performers' training (on tonal works—apparently the software industry is not the only realm plagued with backwards compatibility!). The vertical extent of the staff comes historically from

[86] In recent decades any generalization claimed about music is met with vociferous exceptions. The present discourse, for clarity and finitude, occasionally generalizes outrageously like this.

vocal range. Clefs, ledger lines and the grand staff extend its usefulness beyond what a tenth-century monastic choir could handle. But this notation, this language, still suggests that an instrument is meant to stay in one limited tessitura. Finally, what of the staff's horizontal structure? Bar lines imply a (possibly irregular) pulse; the predominantly binary relation between note durations also implies pulse. But this horizontal structure is the weakest aspect of the staff, most frequently modified even from its earliest history. This is also reflected in the construction of orchestral instruments, which can far more easily play irregular rhythms than irregular pitches (microtones).

Dynamic markings—*piano*, *forte*, "hairpins"—historically developed much later than markings for pitch and duration. Italian words spelled out in full came first. These became impractical as they occurred more often (because they grew to have greater compositional importance, no longer a parameter left to the instrumentalist), so naturally the abbreviations *p* and *f* developed. Two levels of dynamic then were insufficient, and the number slowly grew (Beethoven needed no *ppp*'s and only a few *fff*'s in his entire opus; today an *fff* is a commonplace). The requirement for gradual transitions from one dynamic level to another prompted the hairpin. Some total serialist compositions took this to the apparent extreme, placing separate dynamics on every note in the score; but the true extreme is a separate staff for dynamics, graphing loudness as a function of time equally as important as pitch.

Timbre developed even later than dynamics. Some early instrumental scores have not even an instrumental designation, a puzzle for musicologists of later centuries. *Arco* and *pizzicato* are the best-known early example of notated varying modes of playing; this textual presentation continued for non-*modo ordinario* playing of strings and other instrumental families, abbreviated where necessary.

Italian adjectives certainly describe timbre, but in a larger role they suggest the overall character or affect of some portion of the music. They indicate "performance practice" unformalized modifications of rhythm, dynamics and timbre. They indicate, if anything, the composer's deliberate (and sometimes practical) *lack* of mental model of the instrument. In using these adjectives the composer focuses on the emotional state of the performer and listener, and leaves it to the performer to decide what means are employed to the desired emotional end. He has a mental model of the performer, not the instrument.

2.3.5 The desirable difficulty of learning

We can desire three things from a synthetic instrument: (i) a rich variety of sound, yet with subtle nuances; (ii) many dimensions of control; (iii) fairly simple mappings between them, so the performer thinks about *functional* rather than *operational* tasks: what perceptual result to produce, not what mechanical actions are required to obtain that end (Garnett and Goudeseune 1999).

Musical instruments seem to be inherently difficult to play; simplifying this task may not be desirable (Trueman and Cook 1999; Ungvary and Vertegaal 1999). This difficulty is not something to be avoided and hidden. Folk engineering wisdom says that if you build something that even an idiot can use, only an idiot will want to use it. More formally, a hidden cost often comes with designing an instrument on which novices quickly achieve gratifying results: later progress may be more difficult than it would have been with a less accommodating interface. While convenient for writing business letters and reports, the increased automation of word processors is a poor model for a composition tool: one should rather *problematize* such a user interface (Hamman 1997b).[87] This example applies to our field too: "Though the principle of effortlessness may guide good word processor design, it has no comparable utility in the design of a musical instrument" (Ryan 1992, 416). The positive value of effort is shown in the power and drama of a solo piano performance of Mussorgsky's *Pictures at an Exhibition* when compared with one of its orchestrations (or, more extremely, with a recording). It is no great feat for an orchestra to make wide leaps of pitch or sudden changes of dynamics, even less so for a compact disc player. But instruments need not be designed with deliberate difficulties; it suffices to remember Einstein's dictum that things should be made as simple as possible, but no simpler. Ease of use has questionable value when obtained at the expense of expressive power and flexibility.

At the professional level all orchestral instruments can be said to be about equally difficult to play well. Particular techniques are easier on certain instruments. Performance technique and mechanical developments have led each orchestral instrument to a point where it takes a performer's complete cognitive (never mind emotional and spiritual) ability to play it well. (In extreme cases, say the triangle, the performer plays other instruments also; but this exception proves the rule, which is fundamentally one of economics.) If we then extend such an instrument and demand more of the performer with these extensions, we must demand less in other areas. For example, the eviolinist gives up virtuoso left-hand technique to attend more to spatial motion.

At the low level of operational tasks, performers remember two classes of things: methods (performance heuristics) and artifacts (static elements like grip, hand shape, embouchure). Artifacts start as primitive operations on the instrument, and become fancier with experience: the pianist learns a note, a chord, a cadential pattern, the chord changes to *Autumn Leaves*. Aspects of this learning and memory include: muscle memory; matching a remembered timbre (if the timbre space is continuous and hence

[87] Word processors correct our typing mistakes as we type them, one-button cameras set focus, aperture, exposure for us, car doors lock without our explicit bidding. A system like this which "does things for us" necessarily assumes more about the user; this is perilous for any instrument intended for more than one composition. Such an instrument may be harder to play since the performer must internally model its helpful behavior.

navigable); sequential memory; and transfer from short-term memory to long-term memory. This last one works by grouping low-level units into higher ones.[88] Long-term memory builds up a codebook-like structure identifying similar (low-level) patterns and recalling each instance as a variation on the basic model. The raw data is stored as a set of generative elements and their transformations, saving memory and showing links in the data.

At the highest level of functional tasks, performers wants to express something through playing a composition. What is expressed is of course not verbal: in that case a leaflet instead of a performance would suffice. But if there are performers, they presumably want to contribute some hard-to-quantify thing to the composition. In common-practice music, most studies of musical expressiveness (such as those in the journal *Music Perception*) consider how performers depart from metronomic and dynamic precision in a plain score to meet certain goals, typically using adjectives like bright, mellow, somber, agitated. These studies also consider how listeners evaluate different performances of a single score with such adjectives. This adjective-based approach begins to quantify how good a performance is: How technically unflawed was it? and did it meet the adjectives I had in mind? This approach also applies to more recent scores insofar as the performers still contribute something to the music.

[88] This probably underlies Lerdahl and Jackendoff's generative theory of tonal music (Lerdahl and Jackendoff 1983, 332).

3. Multidimensional Control

We have seen how a wide range of musical instruments convert gestures into sound, how an instrument can be broken down into individual controls driving individual acoustic dimensions (input parameters and output parameters), and how composers have addressed the problem of creating music with regard to individual parameters. Now we are in a position to speak of the general problem of multidimensional control from a mathematical rather than a musical point of view. The previous two chapters have investigated music and constructed a language of control theory around it; in this chapter we turn around and play the role of the control theorist investigating music. The general results we find are then applied to design a synthetic instrument, the eviolin, in the subsequent chapter.

In short, this chapter considers how to control a real-time process which takes as input a fixed set of continuous and discrete parameters. We are particularly interested in reducing the complexity of a large set of parameters so a single person can control it. The output of the process is usually acoustic, but it could equally well be visual, chemical, or abstract. As context requires, therefore, we interchangeably refer to the *user* or the *performer* of the controller.

Parameters which take on a discrete set of values can generally be reduced to continuous parameters. If the values can be ordered from least to greatest, the parameter can directly be treated as continuous though with coarse resolution. If the values are not so orderable, as with the *ad hoc* timbres offered by the stops of a pipe organ, they may be embedded in a space of dimension greater than one by means of perceptual discrimination, situating similar values close together. (This is attempted in two dimensions by the layout of both organ stops on a console and orchestral musicians on a stage.) Some comparison (*i.e.*, ordering) of discrete values is necessary for any generalization or theory. Since such an ordering leads directly to speaking of points in a continuous space, we shall principally deal with parameters which are continuous.

This chapter uses some technical language. Besides drawing on the previous chapter's terminology—controls of certain orders driving dimensions with some spatial and temporal resolution—extramusical fields of knowledge are referred to such as algebra, Euclidean geometry, the analysis of algorithms (big-oh notation), interpolation theory, psychoacoustics, neural networks, and algebraic topology. This chapter would be prohibitively lengthened by including all these prerequisites, so instead it is prefaced with an overview and summary of results. Some terms mentioned in this overview are expanded only in the main part of the chapter.

Figure 14 summarizes a simple taxonomy of input devices for musical instruments, and highlights which devices offer a large number of degrees of freedom without imposing high selection overhead (the effort required to begin using an input device). Given these input devices or controls, we then evaluate ways to construct a smooth mapping from the several degrees of freedom (or dimensions, geometrically speaking) offered by the controls of an instrument, to a possibly different number of perceptual dimensions associated with its sound. As a musically useful starting point, such a mapping can be built up from a *pointwise map*, an association of particular input values with output values: when the performer does *this*, the instrument should sound like *this*. The job of an *interpolator* is to then produce reasonable intermediate outputs for intermediate inputs. Scientists have invented many techniques of interpolation to generate intermediate data from discretely sampled data, such as inches of rainfall, pollution levels, or simply elevation. My own technique, simplicial interpolation, works better than existing ones in the application domain of musical instruments. Only one technique, natural neighbors interpolation, is as good in all respects as simplicial interpolation while being smoother, but it is quite difficult to implement for more than two dimensions. Therefore simplicial interpolation is used to map the various input dimensions of the eviolin to its acoustic output dimensions.

The pointwise map which is given to the interpolator can be defined automatically instead of manually. The synthesis component of an instrument can be automatically "played" and explored to produce a small set of input points (sets of values which the inputs have), so that the points correspond to sounds which are perceptually distant from each other. This set of points compactly represents the entire gamut of timbres available from the instrument. This automaton, the *timbre rover*, has generated many timbre spaces for the eviolin.

Next we consider how controls can be combined to work together, and review experimentally measured limits of human attention, speed, and accuracy in performing a task. These limits naturally apply to a performer manipulating the several controls of an instrument.

Lessons from the field of virtual environments are applied to synthetic musical instruments. Designers of virtual environments and of synthetic instruments share many challenges. Good perceptual-motor coupling or "feel" requires appropriate choices for the gain of each control, various system latencies, intentional hysteresis, and effectively presented visual or haptic feedback. The difficulties of mode in virtual environments suggest that synthetic instruments should avoid many secondary controls: such impose a cognitive overhead of not only selection but also of remembering what is currently selected.

Finally, we compare technologies for measuring the position of parts of the instrument or the performer. These technologies allow directly spatial gestures to be made, and even parsed to simulate many other input devices. Of all systems which are accurate and robust enough for on-stage use, magnetic tracking, particularly the Ascension SpacePad product, is found to be the best compromise between price ($1400) and features (two sensors measuring position and orientation). An efficient calibration procedure for this tracker lets it be quickly set up before a performance.

A note on terminology: by the integers d and e we denote the number of dimensions of the controller's input and output respectively. So the controller implements a mapping from $\mathbf{R}^d$ to $\mathbf{R}^e$.[89] Typically d is 2 or 3, and e lies between 5 and 50; but the mathematics holds for any $0 < d < e$.

3.1 Physical controllers

Recall that by *controller* we mean the complete interface or set of commands by which the user affects the real-time process. This is in the spirit of popular MIDI terms like keyboard controller and guitar controller. By *control*, again, we mean a single indivisible part of the controller.[90]

If a controller has several controls, particular controls can be *selected*, *adjusted*, and *deselected*. Selection is the allocation of performer resources to a particular control; adjustment is the actual change of state of the control; and deselection is the relinquishing of resources in preparation for subsequent selections.[91] Less abstractly, when a performer selects a control, he directs attention and possibly executes some muscular motion. When the muscular motion involved in adjusting a control is much more prominent than that involved in selecting it, the latter may appear to be implicit. Such is the case with singing: a singer wishing to adjust pitch, vowel color, or loudness does not apply certain muscles before the fact; he simply attends to the control needing adjustment, a motion of attention rather than a motion of muscles. Thereupon he can immediately adjust the required controls (while trying to not overmuch vary the other controls less strongly attended to). When sufficient resources are available to adjust all controls at once, selection is superfluous.[92] It can then be considered to occur just once, when the instrument is picked up.

---

[89] This customary notation of computational geometry causes no ambiguity with the other meaning of *e* as the base of the natural logarithm.

[90] We do not here consider controls with temporal behavior, such as auto-repeat keyboards or touch-plate light switches which turn a light on or off or vary its brightness as the plate is tapped or held in various ways. Such a control can have a vast range of possible behaviors; its produced gestures can be oddly constrained since the parameters driven by it are not free to be varied directly.

[91] By resources we mean the performer's finite cognition, attention, and physique: only this many limbs with this range of motion, this spatial resolution, and this speed.

[92] In particular, a controller which consists of only one control needs no selection mechanism. The bugle exemplifies this: it always has *some* lip pressure. Lip pressure needs no preparatory physical motion before it can be adjusted.

Selecting a control may be as simple as moving a limb to it. But sometimes it may be unwieldy to have all the controls directly accessible by mere positioning of a hand. Then we need *secondary controls* to change the behavior of the *primary controls*, like how the caps-lock key on a typewriter keyboard changes the results of pressing the alphabetic keys. Indeed, the various "shift" keys on computer keyboards (seven on the keyboard I write this with) increase the variety of commands available from more basic mouse and keyboard gestures without unmanageably increasing the size of the keyboard. A single multi-way selector switch can also associate physical gestures performed on a primary control with the changing of value of one of several sets of scalars.[93] It is difficult to rigorously define when a control is secondary, but the label applies well when several of the following hold: the control modifies the behavior of another control which is manipulated more often; the control can be left unattended for a while, physically or only attentionally; the control causes a sound or behavior which is in some way nonstandard.

Introducing secondary controls reduces the number of primary controls, thereby simplifying the controller.[94] Of course this is a compromise: simultaneous adjustment of multiple controls is reduced. Also, the interface is deepened even while it is narrowed: the performer's mental model of the instrument is more elaborate and takes longer to learn. A range of compromise in fact exists. At one extreme there are k primary controls (strings on a stringed instrument, keys on a multiply touch-sensitive clavier) and no secondary controls, at the other extreme one primary control with a single k-way selector switch. Between these two extremes there may be m primary controls with an n-way selector switch, where $mn \geq k$. If only a few secondary controls extend the interface of an orchestral instrument, they can often be operated by the feet, for instance as a bank of toggle switches or a three- or four-way "gas pedal."

Secondary "set-and-forget" controls (for timbre, typically) are invoked to change an internal state of the instrument and then left alone for a while. Familiar examples of this are "stomp box" effects pedals for electric guitarists. Of course the nature of the instrument changes with virtuoso use of such controls, as with Vinko Globokar switching trombone mutes every few seconds or Christopher Bockmann playing the organ stops as much as the manuals. Such playing techniques show that there is actually a continuum between secondary and primary controls. Where a control lies on this continuum depends on how flexible the gestures performable on it are, and how awkward it is to select and deselect the control.

[93] Such associating is sometimes called patching. A secondary control could patch the primary control simultaneously with several sets of scalars. The resulting gestural constraint between these sets of scalars would in general be awkward, though.

[94] The literature of process control calls this time multiplexing, where a primary control drives different parameters at different times. Space multiplexing, on the other hand, means one primary (scalar) control per scalar parameter (Buxton 1986).

3.1.1 Sliders

By the term *slider* we mean a continuous control. We distinguish one-dimensional *scalar sliders* and higher-dimensional *multisliders*. Examples of scalar sliders include linear sliders and rotary knobs in a physical control apparatus, and sliders or scrollbars on a computer display. Sliders can of course be nonmanually controlled, as is possible with force sensors, proximity sensors, and thermometers. Joysticks, mice, and virtual reality “wands” are examples of multisliders. As a multislider concurrently drives several dimensions, it is suitable when performance gestures are desired more in the product space of these dimensions than in each individual dimension. (To draw freehand, one prefers a mouse; to draw straight vertical and horizontal lines, the humble Etch-a-Sketch is better.[95]) Considering which dimensions are coupled in performance gesture and which are independent shows the instrument designer where multisliders are appropriate.

A slider can be *absolute* or *relative*. An absolute slider has a direct correspondence between the slider’s position and the scalar’s value. A relative slider, in combination with a secondary control, can change the origin of its coordinate system relative to that of the scalar. This secondary control is usually a momentary switch. When the switch is held, the slider is active: moving the slider adjust the scalar’s value. When the switch is not held, the slider is inactive and can be moved without adjusting the scalar’s value. Another way to think of this is that the switch changes the slider’s behavior between changing the parameter value and changing the origin of the coordinate system. A more sophisticated kind of relative slider uses another *slider* as its secondary control: the secondary slider varies the gain of the primary slider. An example of this is steering wheel as primary slider, gas pedal as secondary slider. The more the gas pedal is depressed, the greater the response from the steering wheel. To be complete, the stick shift acts as *another* secondary control on the gain of the steering wheel.

A relative slider is useful if the effective range of the slider needs to extend beyond its physical range (“pawing” a computer mouse—here the secondary control is lifting the mouse from the desk).[96] Relative sliders are rare in synthetic instruments and absent in orchestral instruments. The first situation may be because the time required to perform a change of coordinate system undesirably constrains real-time performance, or also because the secondary controls introduce what human-computer interface specialists call *mode*, undesirable state in the interface itself. The second situation is more easily explained by the difficulty of building a relative slider without using data-processing equipment.

---

[95] Buxton (1986) analyzes the gestural strengths of the pen, the Etch-a-Sketch, and a hybrid device called the Skedoodle.

[96] The common computer mouse is a relative slider, considered as a physical manipulandum. But it is an absolute slider, considered as a cursor on a screen. We usually cognize the mouse as cursor, not manipulandum. Some applications (notably video games) use a mouse without a cursor; here the mouse’s relative nature is more apparent.

3.1.1.1 Bank of sliders

Perhaps because it is easy to implement with toolkits for computer graphical user interfaces, a common paradigm for controlling the values of a set of scalars is a collection (or *bank*) of sliders. The user can change one scalar at a time by selecting an individual slider (moving the mouse cursor to it) and then modifying it (clicking and dragging the mouse). Various keyboard commands (arrow keys and the like) may work too, but in general they provide no functionality beyond that of pointing-clicking-dragging. The strength of this controller is its simplicity. Its main drawback is its severely restricted gestural repertoire: only one parameter at a time can be varied, so correlation of parameters is not possible. A bank of sliders is better suited for "set-and-forget" parameters, or for finding particular combinations of parameters to be used in other high-dimensional controllers, than for direct real-time control.

3.1.1.2 Bank of multisliders

On a computer screen a multislider might be displayed as a rectangular region with a cursor somewhere in its interior: this would control two scalars corresponding to the *x*- and *y*-coordinates of the cursor. In a virtual reality environment, a cursor in a cube could control three scalars. Representations beyond the literally Euclidean are possible, too. A vertically drawn scalar slider may react to nonvertical mouse movement as well: horizontal displacement or change of angle of a glyph could control a second scalar, perhaps numerically displayed below the slider. A three-dimensional motion-tracking input device (possibly with a visual representation, a cursor in a cube) could have an orientation of yaw, pitch, and roll as well as a position (*x*-, *y*-, and *z*-coordinates) resulting in simultaneous control of six, not three, scalars.

As with scalar sliders, several multisliders can be combined into a single controller. This slightly generalizes the bank of scalar sliders, allowing for correlation of parameters within a particular multislider. But this expressive restriction is still severe. Consider an instrument with 12 parameters. Ideally all 66 possible pairwise correlations are available for musical expression. A bank of two 6-dimensional sliders provides 30, four 3-dimensional sliders only 12, six 2-dimensional sliders only 6, twelve 1-dimensional sliders naturally zero. Generally, as the number of individual sliders in the bank increases beyond two or three, the gestural flexibility of the controller very quickly decreases.[97]

[97] We mean expressiveness in the information-theoretic sense. Instruments with greater bandwidth (larger simplest description of the data communicated) from performer to instrument, and from instrument to listener, we call more expressive. In any particular musical context, of course, the composer, performer and listener greatly affect the net expressiveness.

Vertegaal, Eaglestone, and Clarke (1994) have investigated the use of several hardware interfaces for exploring a four-dimensional timbre space (overtone content, brightness, articulation of attack, and speed of envelope). (Note that their interfaces are intended more for psychological experiments than for musical performance.) One of these interfaces consists of a single four-dimensional slider, a glove tracked in three-dimensional position and in one rotational dimension (roll). The other interfaces have a single two-dimensional primary control combined with secondary controls to apply it to either the first two or the last two dimensions of the timbre space. This division into two pairs of dimensions is reasonable: the first pair deals with steady-state spectral content, the second with the attack.

The mouse interface used by Vertegaal et al. is conventional. As primary control the mouse directly moves one of two cursors in two square fields. As secondary control the mouse is positioned over one or the other of the two squares; depressing the mouse button converts the mouse motion to the primary control. In the joystick interface, two buttons act as secondary control to direct the joystick motion to the first or second pair of dimensions. Besides these three relative sliders (glove, mouse, joystick), they also tried the joystick as an absolute slider. This may have been more successful without secondary controls, since a button-press attaching the joystick to a particular pair of dimensions would make the values of those dimensions jump discontinuously. Even with only one hand available, a pair of absolute-position joysticks would have worked better than a single absolute-position joystick with two patch buttons.[98]

The glove controller fared most poorly in this study. This is attributed to its high latency, but the subjects may have also had difficulty cognizing three separate presentations of information: glove position; a pair of cursors drawn on a two-dimensional grid; and the actual sound. The visual two-cursor display corresponds easily with the planar control offered by mouse and joystick, less so with the height and roll angle of the glove. We can generalize: in some cases visual feedback may distract rather than assist the performer. This can happen even with a simple joystick if its *x-y* position is rendered indirectly, for instance as the size and hue of a colored disc instead of as the *x-y* position of a point in a square.

3.1.1.3 Bank of sliders with multiple selection

A bank of sliders with multiple selection differs from the previously described bank of sliders only in that several controls may be selected and manipulated simultaneously. This typically requires two-handed or multi-fingered operation and thus needs a hardware interface richer than mere mouse and

[98] Another possible solution is a single joystick with force feedback, pushing the stick to the current position of the corresponding pair of scalars when a patch button is first depressed; but this introduces undesirable latency since it takes some time for the stick to reach that position. A zero-latency solution is given by a relative-slider touch screen.

keyboard. The MIDI *fader pack* used in commercial music production exemplifies this, an array of physical sliders together with software to map the value of each slider to an arbitrary scalar (MIDI channel and controller number) in the music production system.

If all sliders in a bank can be selected simultaneously, the bank differs from a single multislider only in that it provides a set of primary axes for the space in which gestures are performed. "Rotations" of a gesture will be more difficult to perform with an array of sliders than with a true multidimensional slider (think of drawing straight *diagonal* lines with an Etch-a-Sketch).

On the other hand, all controls may not be simultaneously selectable. This may be due to either attentional or physical limitations of the performer. Attentional limitations, as is the case with multiple stopping on a violin, lead to gradually more constrained gestures as the number of selected controls increases. Physical limitations, as is the case with a pianist's ten fingers and finite hand span, impose hard rather than gradual constraints on which gestures are performable. This is because of rehearsal: the violinist can learn to play double stops with more accurate intonation and bowing, in a trade-off between accuracy and rehearsal time. But no amount of rehearsing will cause a pianist to grow extra fingers.

A study more recent than that of Vertegaal et al. found that a multislider sometimes fared better than a bank of sliders with multiple selection, even when the subjects could not verbalize how the multislider worked. Hunt and Kirk (1999) conducted several thousand trials where subjects used three different input devices, a mouse controlling four sliders drawn on a screen, four physical sliders, or the combination of a mouse and two physical sliders (which we will call a single multislider by a slight abuse of language).[99] With each of these input devices the task was to duplicate a short sound presented to the subject; the sound might vary in volume, pitch, unidimensional timbre, or stereo position. In all cases subjects performed better with the bank of sliders with multiple selection than with the mouse, as predicted by our theory of selection overhead. In simple tasks where only one parameter of the sound changed, the bank of sliders with multiple selection showed the best performance though the multislider showed improvement as trials progressed. For complex tasks the multislider was best. This is remarkable when its design is examined. The four parameters were not simply assigned to slider 1, slider 2, mouse x-position, and mouse y-position. Rather, each parameter depended on expressions like overall mouse speed plus average of the slider positions, or mouse y-position plus the speed of slider 1. The experimenters conclude that the simplest mapping of controls to synthesis parameters is not necessarily optimal. (We look forward to another study where the four parameters *are* simply assigned to the four obvious linear controls, for an even stronger conclusion.)

3.1.1.4 Bank of multisliders with multiple selection

An immediate example of a bank of multisliders with multiple selection is a bank of joysticks. More commonly, handheld video game controllers have several thumb-operated joysticks or eight-way hat switches. These small banks are typically controlled manually, one hand per multislider; larger banks of multisliders may be rare because they tax the attention of a performer. Perhaps the largest practical banks are motion-capture devices which track a dancer's body. Several dozen three-dimensional positions and/or joint angles can be captured by the best systems currently available, with latency under 5 msec. Dancers well trained in isolation exercises appear to control a few dozen parameters independently. (There have been no formal studies measuring this number, perhaps because in practice it is difficult to adequately define these parameters and determine their level of conscious control, perhaps because the military has had more funding for human factors research than have dance departments.)

The Continuum controller is a sophisticated multislider with multiple selection. This keyboard-like instrument is about as large as a piano keyboard but has a smooth surface. It tracks the position of up to ten fingers on this surface with a spatial accuracy of 1 mm, as well as the downward force of each finger; latency is about 4 msec (Haken et al. 1997). Multiple-selection controllers may have an additional implicit secondary control. Each multislider in the bank is enabled by depressing a momentary switch (touching the Continuum's surface). This is particularly natural for polyphonic controllers, where each additional selected control produces an additional sound.

3.1.2 Common physical input devices

Physical switches can be momentary as with a computer keyboard, or latched like a ballpoint pen or toggle switch. They may provide feedback visually (lights), acoustically (key clicks) or haptically (electromechanical or piezoelectric vibrators).

Multi-way selector switches can be used to "patch" or associate physical gestures performed on primary controls to the changing of value of one of several sets of scalars. A *k*-way selector switch can be

[99] Technically it is a bank of two scalar sliders combined with a two-dimensional multislider, but as both hands can easily operate all the controls at once we can consider it a single four-dimensional multislider. Even the bank of four scalar sliders could be called a multislider. Choosing which scalars to group into a vector depends on which scalars work together in performed gestures. In this case all scalars work together in Hunt and Kirk's experiments, so calling it a single multislider is justified.

implemented as a single switch (like the rotary channel selector of an old-fashioned television), or as an array of $\log_2(k)$ binary switches, or intermediately as $\log_n(k)$ n-way switches or as a nonhomogeneous collection of switches. These switches are usually momentary and fingered. The physical layout of a single selector switch also affects its gestural repertoire. We note particularly that the loop (the television example above) and the line enforce an order of visiting states of the switch, whereas the star topology of an automobile gearshift allows transitions directly between any state and a special nil state. The joystick "hat switch" combines these two: transitions between all states and the central nil state are possible, while rolling the thumb in a circle produces the state transitions of a loop. A single switch allowing all possible state transitions is physically difficult to build; several switches are preferred if more transitions are required than loops, lines, and stars.

Rotary knobs and linear sliders provide inexpensive control of continuous parameters. Bounded sliders typically work absolutely while unbounded sliders are naturally relative; the latter usually drive an optical motion sensor to allow for mechanical decoupling. Unbounded sliders, often rubber bands or treadmills mounted on two rollers, usually need explicit visual feedback, often a strip or ring of LEDs mounted in or beside the device. An array of knobs, sliders, or switches is sometimes built into a single unit, as with MIDI fader packs or dial-and-button boxes.

The simple joystick is like a pair of rotary knobs cross-coupled. Some joysticks can be twisted as well, offering a third cross-coupled knob. Joysticks may remain stationary when released; *auto-centering* joysticks return to a central position when released. (Reduced cross-coupling occurs when only some of the joystick's degrees of freedom are auto-centering.) *Force-feedback* joysticks generalize the simple spring behavior of self-centering: they offer variable resistance to motion or haptic display of various impulses and vibrations. The *PHANTOM* is an articulated arm which senses position, and applies force in, three dimensions. It has a range of 40×60×80 cm with 20 μm accuracy, 150 g inertia, and 20 N peak force (SensAble 2000); it has been extended to sense orientation (.01 degree resolution) and apply torque (peak 0.6 N·m), for an impressive total of 6 dimensions each of input and output (Chen 1999). The *SPIDAR* force feedback system senses position of, and applies force in three dimensions to, a manipulandum held in place by taut wires (Cai et al. 1995); similar wire-based systems exist as well (Agronin 1987). Two SPIDARs or PHANTOMs can occupy the same space, for limited two-handed manipulation.[100]

[100] Programmable force-feedback devices clearly demonstrates that an input device can also be an output device, but this duality is true to some extent of all nonstationary input devices. Even a simple mechanical switch offers haptic feedback.

Pressure sensors, commonly force-sensing resistors (FSR's), offer continuous control in one dimension. They can be combined into an isometric joystick (two degrees of freedom, three if twist is also measured) or a *spaceball* (six degrees of freedom, strongly cross-coupled; typical latency 40 msec (Spaceball Technologies Inc. 1993)). Pressure sensors are readily applicable to nonmanual input: breath pressure, bite pressure, weight distribution on feet or chair, clavier aftertouch.

The *ribbon controller* and *touchpad* track the position of a single fingertip along a linear continuum and on a surface respectively. Advanced models also measure the pressure applied by the fingertip; Tactex Controls (1999) even sells a touchpad which tracks several fingertips at once (8-bit pressure resolution, spatial resolution 1 mm, latency 5 msec). A touchpad may be divided into several virtual devices, an arrangement of sliders, momentary switches and smaller touchpads on a single physical surface. This is the input-side analog to a window manager for graphical output (Buxton, Hill, and Rowley 1985). Despite offering less proprioceptive feedback, relative devices often work as well as absolute ones; for example, the touchpad pointing devices on laptop computers are relative. Rudimentary tactile feedback can be provided for multiple virtual devices by laying a cardboard template on the physical device (Buxton, Hill, and Rowley 1985). Brown, Buxton and Murtagh (1990) show with several examples that such feedback often makes looking at the devices unnecessary while operating them; this is of value to musicians already looking at scores, conductors, or other musicians. On the other hand, a need for complex visual feedback may justify the extra expense of a computer display integrated into the touchpad.

Light pens and *tablets* with styli are related to touchpads. They are more accurate than touchpads but require the hand to hold a stylus, so they work poorly when the hand also has other tasks to perform and must therefore grasp and release the stylus. Several buttons may be mounted on the stylus, offering a greater repertoire of gestures than a touchpad.[101] The common computer mouse is analogous to a stylus without a tablet. The common trackball, another two-dimensional continuous controller, has the interesting property of inertia: the hand can set it spinning and then proceed to another task while the ball is still rotating. But this is better suited to pursuit tasks than to precisely locating a cursor.[102]

Several technologies exist for remotely measuring the position and/or orientation of a sensor freely moving in space. We consider their individual merits later; here it suffices to say that such a sensor is the

[101] The gestural repertoire of a tablet sensing pressure and angle is remarkably similar to that of a bowed string (Serafin et al. 1999).

[102] Similar inertia is simulated for a mouse by the XWindows program *xnuisance*; it is impressively annoying. Unbounded linear sliders or knobs with inertia are possible, for instance the heavy tuning dial on some radios in the 1970's.

three- (or six-) dimensional version of a tablet plus stylus. The working area of the tablet is a bounded volume instead of a surface. Instead of varying the pressure between stylus and tablet surface, the user can push on a pressure sensor attached to the position sensor, as with the four-way pressure sensor mounted as a thumb-controlled joystick on the wand used in the CAVE virtual environment. (Several momentary switches are also mounted on the wand.) Latency of motion trackers is moderate for musical applications, typically 15 to 50 msec; a common bottleneck for latency in motion tracking is, ironically, a slow RS-232 serial connection between tracker and computer. Most of these technologies can track multiple sensors in the same volume, a notable difference from most tablets and ribbon controllers.

Many other specialized sensors have been used as controls: eye trackers, exoskeletons, force-feedback mice, steering wheels and yokes (qualitatively different from knobs and sliders in the muscle groups used), and sensors for muscle contraction (EMG), brain wave activity (EEG), illumination, proximity, temperature, voltage, and so on.

Figure 14 tabulates the largest family of controls, those manipulated by a hand or finger. It lists only nondecomposable controls; compound controls like a wand with buttons or MIT's Digital Baton can be reduced to the elementary units listed here.

Selection overhead → ***Type of measurement*** ↓	*low*	*medium*	*high*
Discrete state	momentary switch, toggle switch, hat switch	multi-way switch	numerical keypad
Fingertip location (and unidirectional force)	ribbon (1), touchpad (2)	multi-touch pad, Continuum (3–15)	
Unidirectional force applied by hand	force-sensing resistor (1)	torque sensor (1), isometric joystick (2–3), spaceball (6)	
Location (and orientation) of manipulandum	slider (1), large knob (1), trackball (2)	small knob (1), mouse (2), joystick (2–3)	tablet+stylus (2–4), wand (3–6)
Location (and orientation) of, and force applied by, manipulandum			pressure-sensitive tablet+stylus (3–5), force-sensing mouse (4), PHANTOM (6), SPIDAR (6–9)[103]

Figure 14. Elementary manual controls. Parenthesized numbers indicate how many degrees of freedom a control has.

Beyond what is shown in figure 14, recall that a control may be:

- primary / secondary;
- separate / an element of a bank / an element of a multiply selectable bank;
- absolute / relative;
- bounded / unbounded in its motion;
- with / without visual feedback (at the control itself, or separately in a computer display).

Figure 14 shows that selection overhead generally correlates with number of degrees of freedom. Efficient controls, those with a high ratio of degrees of freedom to selection overhead, are: touchpad, trackball, multi-touch pad; spaceball; orientation-tracked SPIDAR. Other considerations being equal, these are therefore especially recommended as manual controls for synthetic musical instruments. Those with low ratios may have other advantages like high resolution or small size.

[103] A SPIDAR-like system could be built with torque sensors on the manipulandum, for 12 degrees of freedom: 3 each for spatial position, spatial orientation, translational force, and rotational torque. The challenge in building such a device is keeping the inertia low.

3.1.3 Muted orchestral instruments

Orchestral instruments can be themselves used as controllers. The sound such an instrument produces can be analyzed or *tracked* into parameter streams which then drive sound synthesizers. Examples of such parameters are pitch, amplitude, and timbral information amenable to real-time analysis such as spectral centroid and amount of unpitched noise. The literature speaks of *pitch tracking* and *amplitude following* or *envelope following*. Artificially derived data streams like depth of vibrato can also be generated by the tracking system. Since the orchestral instrument itself is now a sensor of gestures and not also a producer of sound, it is appropriate to extend what it senses by adding other devices to it. In particular, extra parameters can be provided by tracking the spatial position and orientation of the instrument; also, the feet of most orchestral performers are available to operate pedals as secondary controls.

If the instrument is muted (caused to produce little or no sound by itself), description and analysis of the resulting sound need not deal with the complication of two separate sound production mechanisms. We consider only the muted case here. Only chordophones are easily muted; the vibrating body in brass, woodwinds, and other aerophones is an air column, while in membranophones and idiophones the vibrating body is typically large and not easily damped. Very quiet percussion instruments are perhaps the only companions to chordophones in having an inherently quiet primary vibrating body. However, thumb pianos, rattles, John Cage's amplified cactus (the spines are plucked) and other members of this family generally have less expressive potential than most chordophones. Of course mechanical claviers can easily be separated from their sound-production apparatus; electric pianos and Hammond organ are common examples. The commercial MIDI synthesizer industry has thoroughly developed this area; along with electric stringed instruments, such controllers are covered in the previous chapter.

3.2 Interpolation

3.2.1 Classical interpolation

In the most general sense, *interpolation* is "the performance of a numerical procedure that generates an estimate of functional dependence at a particular location, based upon knowledge of the functional dependence at some surrounding locations" (Watson 1992, 101). *Extrapolation* is commonly used when the particular location lies outside the convex hull of the (no longer) surrounding locations; we include this case in the term interpolation. The goodness of an interpolation function is described by expressions like maximum error or root-mean-square pointwise error.

We use interpolation as a convenient starting point to construct a mapping between two Euclidean spaces, from a space of dimension equal to the number of degrees of freedom of the continuous controls which the instrument uses as input, to a space of dimension equal to the number of perceptual parameters of the instrument's sound output which we wish to consider.[104]

By the term *synthesis algorithm* we mean a computation (not literally a terminating algorithm) which (i) runs in real time, (ii) produces as output a time series of numbers, interpretable as an acoustic waveform, and (iii) accepts values for a fixed set of continuous- and discrete-valued input variables. (Here we will consider discrete-valued inputs to be approximable by equivalent continuous inputs.)

It is easy to specify several pairs of points in the input and output spaces, in effect specifying "when the controls have these values, make this sound"; interpolation then defines intermediate sounds for intermediate values of the controls. (In fact these pairs can be computed automatically by means of a model of human hearing, as we shall see.) Having tried out the instrument resulting from this map, with more effort one can refine the map by moving input points (make that sound over here instead), moving output points (no, that sound needs to be more like this one), or introducing new pairs of points (adjust the sounds which the interpolator happened to produce in this little region). With an appropriate choice of interpolator we need assume little else about the system. In particular we need not assume linearity of the synthesizer or of human perception. In short, constructing continuous maps by extending pointwise maps via interpolation has the advantages of low effort, generality, scalability and local adjustment.

In this context, we define the unidimensional interpolation problem thus: given a finite pointwise map $S = \{(x_i, y_i)\} \subset \mathbf{R}\times\mathbf{R}$, construct a function $f : \mathbf{R}\to\mathbf{R}$ such that $y_i = f(x_i)$ for all i, and such that f has nice properties. Among these properties continuity is usually most desirable; others are differentiability, having bounded higher derivatives, being C^∞ and having adjustable smoothness and tension. More generally: given sets $\mathbf{A}$ and $\mathbf{B}$ and a function $g : \mathbf{A}\to\mathbf{B}$, construct a function $f : \mathbf{A}\to\mathbf{B}$ such that $f \supset g$ (considering f and g as subsets of $\mathbf{A}\times\mathbf{B}$) and f is nice. Our niceness is not treated directly in classical interpolation theory, because our context has no ideal function which the interpolation tries to approximate; there is no error measurement to speak of. Two desiderata in (Watson 1992, 103) apply, though: the interpolated value should depend only on nearby values (which ones defined as nearby depending only on the spatial arrangement of the points); and the tautness of the map should be adjustable independent of spatial arrangement and of data values. A requirement for us, not a desideratum, is that the

[104] In a more restricted sense, it is useful in instrument design to construct maps from the space of controls to the space of parameters required by the synthesizer (rather than all the way to perceptual parameters).

interpolator work with arbitrary ("irregular") spatial arrangements of points, not only rectilinear grids: without this, the advantages outlined above vanish.

Interpolation theory implicitly assumes certain things which may not hold for musical instruments. For interpolation to make sense, the input and output spaces must themselves be continuous. For the mapping to be repeatable, the input and output spaces should not change with time; in particular, large hysteresis in the sound synthesizer (when its output depends on past as well as current inputs) is incompatible with this method of constructing mappings.

The property that the extended map f agrees with the pointwise map S actually applies only to the class of interpolators called *exact*. Inexact or *approximate* interpolators, notably some spline methods, may be preferred in some cases such as sampled meteorological data or noisy input data. We implicitly assume exactness from now on. A general specification of an exact interpolator, derived from (Shepard 1968), defines f in terms of weighting functions w_i:

$$f(x) = \sum_i w_i(x)\, y_i$$

where

$$\forall i,\ w_i(x_i) = 1; \quad \forall i, \forall j \neq i,\ w_i(x_j) = 0; \quad \forall x,\ w_i(x) \geq 0 \text{ and } \sum_i w_i(x) = 1.$$

This reduces the problem of defining f to defining the weighting functions w_i which obey these constraints. Classical exact interpolators include proximal interpolators, B-splines and kriging (Collins and Bolstad 1996; Bryan 1999).

In *proximal interpolation*, $f(x)$ takes the value of $f(x_i)$ for the x_i nearest x: f is extended constantly from each input point to that point's Voronoi cell.[105] This is inexpensive and predictable, but inappropriate for most musical uses since f is discontinuous. *Inverse distance weighting* (IDW) generalizes proximal interpolation by allowing other input points x_i near x (not just the nearest one) to affect the value of $f(x)$; the weight of an input point is usually inversely proportional to some power of the distance from x to that point. IDW produces smaller discontinuities but does not eliminate them; it also produces undesirable local extrema at the input points. *Optimal IDW*, a variant which tries to minimize mean absolute error, may be desirable where continuity is less important than algorithmic simplicity.

B-splining constructs f from patches of polynomials, so f is not only continuous but C^2. It produces artifacts such as extrema beyond the input values ("overshoots") when the space containing the input

[105] The *Voronoi cell* of a point is all points closer to it than to any other point of this finite set. The *Voronoi diagram* of the set is the collection of all Voronoi cells. Other names have been given to this rather general concept by geologists, botanists, meteorologists and probably others; here we prefer the geometrical name.

points is not smooth, because it prefers minimum curvature to the avoidance of such artifacts. It does best when the input points lie on or near a grid. Cubic splining of irregularly spaced data is not recommended (Eckstein 1989; Hutchinson and Gessler 1994). As musical applications need not in general be smooth (the breaks in woodwind register, for example), B-splining may work well in some cases but it seems dangerous to use it for a general implementation of interpolation. The sophisticated *regularized spline with tension* (RST) avoids overshoots and in fact constructs f to be C^{∞}. Mitasova et al. (1995) describe its implementation in a geographical information system; Mitas and Mitasova (1999) sketch a generalization to more than 3 dimensions, explaining enough to reveal that implementing it takes significant effort.

Kriging considers how quickly the variance between input points changes in various parts of the space (Oliver and Webster 1990; Krige 1981). It is quite effective in geographical applications because it follows anisotropies in the input data and provides error estimates, but in higher dimensions it becomes cumbersome. Sárközy (1998) warns that the data should satisfy several stationarity conditions; these are difficult to assume in musical situations, leading us to *universal kriging*. This adaptation of kriging leads to many possible solutions, which must be narrowed down to one solution by either random or interactive decisions, again undesirable for musical instrument design. Kriging also performs poorly if input data is sparse, which can well be the case with timbre spaces. Finally, (Hardy 1990) demonstrates that kriging is poorer than other methods at handling smoothness, which is an important factor for musical instruments.

Artificial Neural Networks (ANN's) have traditionally been used as classifiers, *i.e.*, deciding which of a set of classes an input point belongs to. But with some work they can be adapted to the task of interpolation, usually as a three-layer feed-forward network (figure 15). The number of neurons in the input and output layers of the ANN is set to be the number of dimensions of the input and output spaces. The pointwise map $\{\boldsymbol{x}_j\}$ is applied to the ANN as training data; the ANN is trained to minimize its error in computing this map, via back propagation. Sárközy (1998) points out that the learning rate of this back propagation is critical to successful training, though some progress is being made to automate the choice of learning rate. One hopes that applying intermediate values to the input neurons will produce intermediate values at the outputs, but it is difficult to analyze properties of this interpolation, and also difficult to get the ANN to compute the pointwise map with zero error (*i.e.*, be an exact interpolator). A particular adaptation of ANN's for interpolation is the use of *radial basis functions* rather than hill-shaped or sigmoidal functions for the neurons in the middle layer (the h_j). Each of these middle-layer neuron corresponds to an element of the pointwise map, and can be thought of as a "bump" with center at $\boldsymbol{x}_j$ and radius determined by how near the other $\boldsymbol{x}_j$'s are. Designing effective multipurpose interpolation tools using ANN's is still an area of active research.

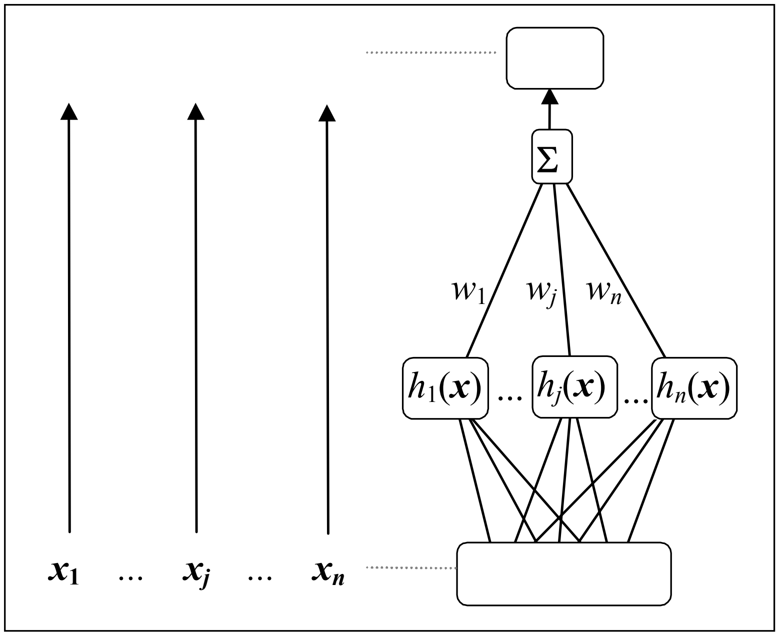


Figure 15. Artificial neural network with radial basis functions, used for interpolation. Three dimensions x_1, x_2, x_3 are input; n dimensions $y_1, ..., y_n$ are output. The pointwise map corresponds to the n training patterns $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$. The $h_j(\boldsymbol{x})$ are the radial basis functions. Diagram based on (Sárközy 1998).

A musical ANN application is found in the expressive real-time speech system GloveTalk II (Fels and Hinton 1998). The position of the right hand in a horizontal plane maps to vowel color. Eleven neurons in the middle layer correspond to eleven basic vowel timbres. The radial basis functions act like Voronoi cells around each vowel point, but with "smooth edges" so dipthongs can be enunciated by sliding the hand from one vowel region into another. The transition regions are quite narrow, though; this stabilizes vowel color but may be inappropriate for musical instruments.

Multiquadrics (Hardy 1990) and polynomial interpolation with basis functions are two examples of at least originally *ad-hoc* function-fitting methods. They have not seen as wide use as the other interpolators surveyed here. This may be because they are difficult to generalize to irregularly spaced data points (Sárközy 1998).

3.2.2 High-dimensional interpolators

"How can I control e parameters with only d scalar controls ($d < e$)?" We can restate this question as "how can I make a *desirable* collection of gestures in $\mathbf{R}^e$ with only d degrees of freedom?" Here $\mathbf{R}^e$ is a family of sounds, perhaps a family of steady-state timbres; what desirability means, we leave undefined in this general context. At any rate, by ruling out undesirable or redundant gestures we hope to find an interesting d-dimensional space inside $\mathbf{R}^e$.

Why is this hope justified? Why do we think it even possible to reduce the number of dimensions from e to d? Stated coarsely, high-dimensional spaces are simply too large to be exhausted by one composition or one instrument. (More technically, the size of the space grows exponentially with the number of dimensions. Problem spaces which grow exponentially—*i.e.*, faster than polynomially—are precisely those which computer scientists call intractable.) Attempting such an exhaustive exploration runs the risk of producing a tedious composition or an incomprehensible instrument. Even in a scientific experiment most of a high-dimensional space is ignored. This premise underlies the field of adaptive mesh refinement, which considers how to best use the finite memory capacity of a computer, and the finite patience of the researcher, to represent and explore only the interesting parts of spaces of merely three dimensions (Bell et al. 1994). So we expect that out of the whole space the subset which we call interesting or desirable—for research, composition, or instrument design—can be adequately described by fewer dimensions than that required for the entire space.

One way to find an interesting d-dimensional space inside $\mathbf{R}^e$ is to first specify several desirable "generating" points in $\mathbf{R}^e$, then construct a d-space enveloping them, and finally define a mapping from the d-space to $\mathbf{R}^e$. If we build up such a mapping from a pointwise map, we call it a *high-dimensional interpolator* or HDI. Beyond the mere existence of this mapping, there are desiderata to balance: continuity, differentiability, linearity, including all important parts of the range space, smooth interpolation between generating points, higher dimensionality of the range space, extensibility to larger spaces, ease of editing the map locally and globally. We begin by requiring two of these desiderata, continuity and smooth interpolation between generating points.

3.2.2.1 Pairwise constraint controller

Consider a rectangular screen on which is drawn a web of a dozen or so line segments (figure 16). The segments meet at labelled points. As you drag one of these points around with the mouse, other points also move around somewhat. If the web controls a sound synthesizer, as these points move around several corresponding parameters of the sound change with them. So this interpolator is not only a formal mapping: its visual feedback and graphical control are essential parts of it.

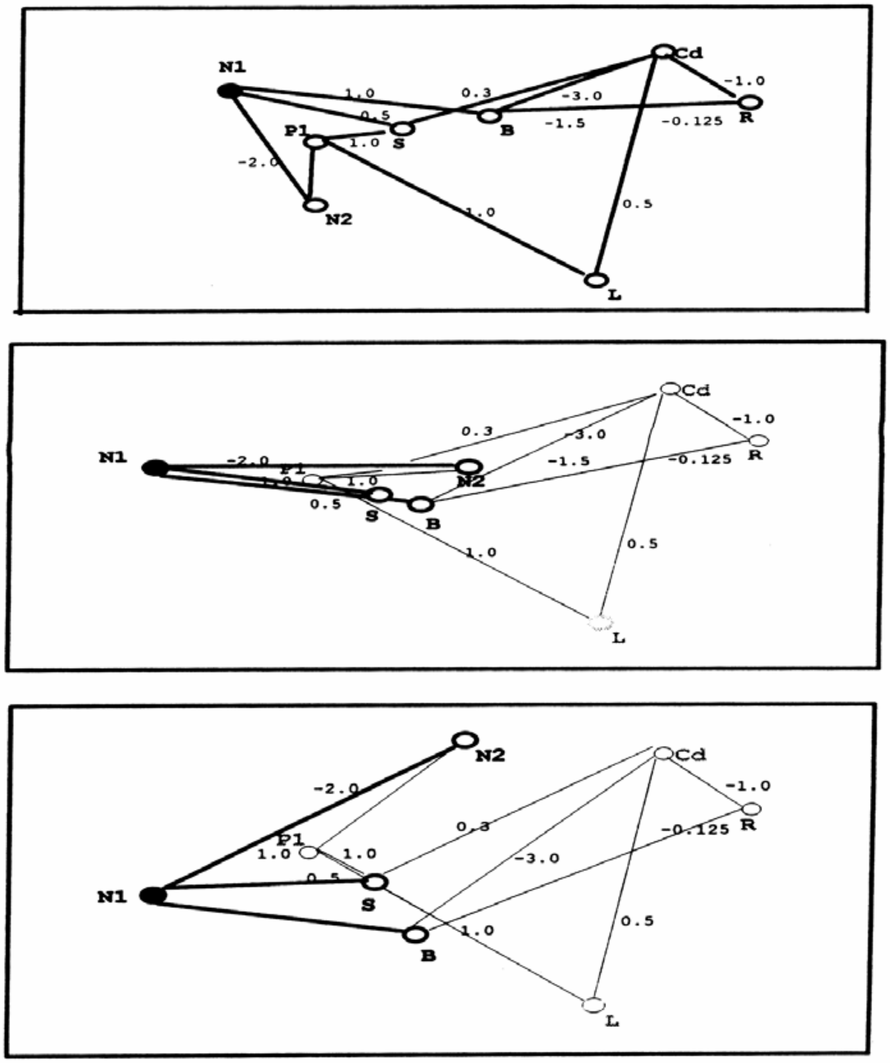


Figure 16. Pairwise constraint controller. As node N1 is moved with the mouse, its adjacent nodes N2, S and B also move. The other nodes, not adjacent to N1, remain fixed while N1 is selected. Reproduced from (Hamman 1997b) with permission.

The pairwise constraint controller maps the positions of e points in a fixed two-dimensional rectangle to one point in $\mathbf{R}^e$. Each of the $\mathbf{R}^e$-point's coordinates corresponds to one of the e points in the rectangle. To each pair of points (i, j) is associated a real-valued weight $w_{i,j}$. A line with a numerical label indicating this weight is drawn between the two points. (Lines of zero weight are not drawn.) The controller is manipulated by clicking on a point with the mouse and then dragging the point around inside the rectangle. As one point moves, the other points are forced to move too. The vector of motion followed by the selected point i is multiplied by $w_{i,j}$ to determine the vector of motion of every other point j. In particular, points connected with zero weight to point i will remain stationary as point i is moved. Also, if the edges with nonzero weight form a tree, then the subtrees from the currently selected point form nested groupings, that is, a hierarchy of control (Hamman 1997a, 112; Hamman 1997b, 106–111).

Each point has two coordinates in the rectangle, but only one coordinate in $\mathbf{R}^e$. This mapping is handled arbitrarily by letting the $\mathbf{R}^e$-coordinate be (a linear function of) the sum of the two coordinates. The most immediate generalization extends the control space from two to three dimensions, letting the points move in a rectangular prism. This gives more room to draw the points and lines, but does not extend the mathematical power of the controller. In fact, if the $\mathbf{R}^e$-coordinate of a point is given by a linear function of the point's coordinates ($x + y$ or $x + y + z$ here), then the control space might as well have only one dimension, a line instead of a rectangle. We prove the two-dimensional case here; generalization to higher dimensions is straightforward. Let the linear function be $f(x, y) = ax + by$ for some constants a and b. If a is zero then we are done, so assume the contrary. Choose any two points i and j, with edge-weight w. Select point i and move it by a vector (dx, dy). Then point j will move with vector $(w \cdot dx, w \cdot dy)$. The i^{th} $\mathbf{R}^e$-coordinate will have changed by $f(x_0+dx, y_0+dy) - f(x_0, y_0) = a \cdot dx + b \cdot dy$; the j^{th} coordinate will have changed by $a(w \cdot dx) + b(w \cdot dy)$. The same effect could have been had by moving the point i by the vector $(dx + (b/a)dy, 0)$: the i^{th} $\mathbf{R}^e$-coordinate would then be $a(dx + (b/a)dy) + b \cdot 0$, the same as before, and the j^{th} coordinate would then be $a(w \cdot (dx + (b/a)dy)) + b \cdot w \cdot 0$, also the same as before. In short, all motions of point i can be replaced by an equivalent motion with an x-component but no y-component, with no change in the behavior of the controller. Since this is true for any points i and j, the rectangle itself can with no change of behavior be projected onto the axis normal to the linear constraint f.

Another straightforward generalization is extending the algorithm to its transitive closure: having moved a point i and its immediately adjacent (nonzero-weight) points j, do the same with each of the points j until all points connected to point i have been visited. But this requires an arbitrary choice of what order to visit the points in, at each stage of the breadth-first traversal. As the connections may not form a tree, this arbitrary choice will affect the behavior of the controller. Also, if many weights have absolute value greater than one, then points far from the selected point i may move very far, even beyond the rectangle. A better approach here is restriction, not generalization: disallow zero weights. The visual representation as it stands becomes cluttered, but all parameters can be controlled simultaneously.

It is initially unclear which subset of $\mathbf{R}^e$ is accessible from a given pairwise constraint controller, since the constraints between dimensions vary. Consider two points i and j with weight 2. If i is selected and moves one inch, j will move two inches in the same direction. But if j is selected, i moves one inch when j moves only half an inch. And if i and j are connected to a third point k, when k moves the relationship between i and j will be different again. It is better to characterize the gestures, local geometric features, which are possible. Each point, when selected, activates a certain set of weights. From the current location in $\mathbf{R}^e$, this set of $(e-1)$ weights defines a vector, a line in $\mathbf{R}^e$ along which the location may move (and thus produce a characteristic gesture, a characteristic set of correlations among the parameters given by the $\mathbf{R}^e$-coordinates). So a pairwise constraint controller could be implemented with only a single scalar controller (slider, data wheel, *etc*., not necessarily bounded like the rectangle is) and an e-way switch, choosing one of e vectors along which to move in $\mathbf{R}^e$. Such an interface would not need a visual component, making it more versatile in musical environments. It also suggests another generalization, having different numbers of vectors than just e. Unlike some other controllers described here, the entire space of $\mathbf{R}^e$ is potentially accessible. This happens if the set of vectors is chosen so as to span $\mathbf{R}^e$; but in this case the controller is functionally identical to a simple bank of linear sliders on $\mathbf{R}^e$ under a change-of-basis transformation.

The value of the original pairwise constraint controller is therefore found not in its expressive power but in its visual presentation,. The displayed points are connected with lines and visibly affect each other, encouraging the performer to consider the input as a whole rather than as its individual parameters; this is paradoxical since the gestural repertoire is the same as a bank of linear sliders.

3.2.2.2 Simplicial interpolator

The simplicial interpolator lets one move a point in a d-dimensional space and thereby control e dimensions at once. It takes as input a set of points in $\mathbf{R}^e$ called *generating points*. Only $d + 1$ points are required, though thousands of points work also.[106] These points are chosen to correspond to *a priori* interesting sounds. From this set of points the simplicial interpolator produces a d-dimensional controller which, moment by moment, takes as input a point in $\mathbf{R}^d$ and produces as output a point in $\mathbf{R}^e$. This output point moves through an unbounded nonlinear e-dimensional region containing the generating points. In a real-time musical application the controller takes as input d real-valued data streams from a mouse, joystick, wand or other device, and outputs e data streams to a sound synthesizer.

[106] At least $d + 1$ generating points are needed in order that the vectors connecting their preimages in $\mathbf{R}^d$ are able to span $\mathbf{R}^d$; otherwise a smaller value of d might as well have been chosen.

The algorithm consists of an initialization phase and a running phase. The initialization phase finds good preimages in $\mathbf{R}^d$ of the input points in $\mathbf{R}^e$, thereby defining a pointwise map. (Good preimage points are those whose pairwise distances approximate the pairwise distances of the input points.) A *Delaunay triangulation* of these preimage points in $\mathbf{R}^d$ is used to extend this pointwise map to a simplicial map defined on the convex hull of the points in $\mathbf{R}^e$. The simplicial map is then extended beyond the hull to all of $\mathbf{R}^e$, to allow arbitrarily large control spaces. In the running phase, the algorithm determines which simplex contains a given *query point* in $\mathbf{R}^d$, computes the query point's *barycentric coordinates* with respect to that simplex, and then finds the corresponding point in the simplicial d-complex in $\mathbf{R}^e$ (figure 17). Two points in $\mathbf{R}^d$ and $\mathbf{R}^e$ correspond if they lie in corresponding simplices, and have the same barycentric coordinates with respect to those simplices.[107]

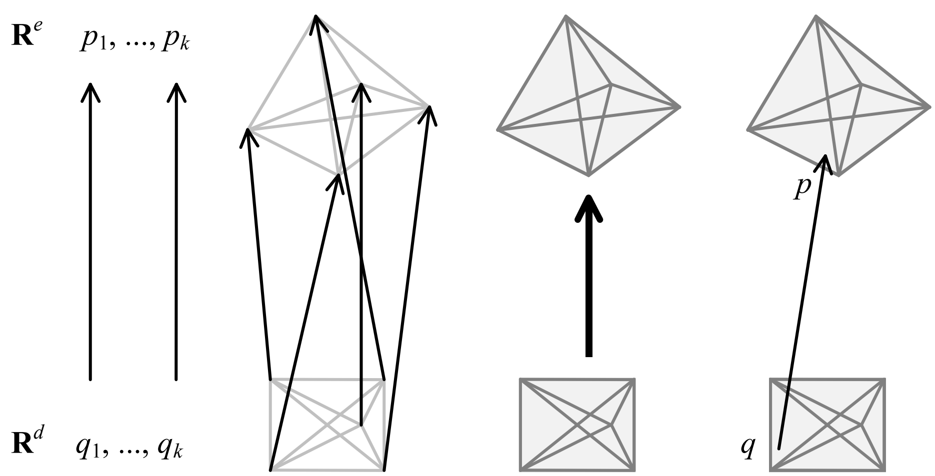


Figure 17. Pointwise map extended to a simplicial map, used to map an arbitrary query point q in $\mathbf{R}^d$ to an image point p in $\mathbf{R}^e$.

The set of k generating points p_I in $\mathbf{R}^e$ induces a set of pairwise Euclidean distances. As a first step in constructing the map from $\mathbf{R}^d$ to $\mathbf{R}^e$, we define preimages q_I of the p_i. Taking the unit d-cube as the domain from which to choose the q_I, we find a set of q_I such that (I) their normalized pairwise Euclidean distances approximate those of the p_I, and (ii) they adequately fill the d-cube. By filling we mean that their convex hull has nonzero volume (to take advantage of all d dimensions at our disposal), and that the projection of this hull onto some principal axis of $\mathbf{R}^d$ yields the unit interval (to take advantage of all the distance at our disposal). Having computed these preimages q_I, the initialization phase next computes a

[107] This approach is due to a refinement by Herbert Edelsbrunner of my earlier algorithm (Choi, Bargar, and Goudeseune 1995).

Delaunay triangulation for them. This triangulation software is based on an efficient implementation by Clarkson, Mehlhorn, and Deidel (1993).

Two methods of choosing the q_I are implemented, a *genetic algorithm* (GA) and *Sammon's mapping* (Kohonen 1997). The population members of the GA are sets of q_I, and the fitness function to minimize is RMS error between corresponding distance pairs of the p_I and a given set of q_i. Sammon's mapping computes such a configuration of q_I by iteratively refining a set of q_I in the manner of simulated annealing. Kohonen (1997, 31–32) suggests that it requires $10^4 \cdot k$ to $10^5 \cdot k$ steps to arrive at an adequate solution, but in fact certain choices of temperature gradient yield a good solution in only $10 \cdot k$ to $100 \cdot k$ steps—if a good result is forthcoming at all. Both Sammon's mapping and the GA are random incremental algorithms. They initially gave poor results, but after observing that most of the improvement happened in the first few iterations, I changed strategy. Instead of running an algorithm once for many iterations, it runs several times for correspondingly fewer iterations. Sammon's mapping uses about 700 runs of $70 \cdot k$ iterations each; the GA uses 10 runs averaging around 1000 iterations each. Both take about 250 msec on current desktop computers to find an adequate solution for moderate values of k. The GA is better at avoiding suboptimal local minima, but Sammon's mapping uses less memory, has better cache usage, is simpler, and certainly scales up better to massively parallel computers. One cannot be universally recommended over the other at this point.

The running phase of the simplicial interpolator begins by inducing a d-triangulation of the p_I in $\mathbf{R}^e$ from the d-triangulation of the q_I in $\mathbf{R}^d$. Given a query point q in $\mathbf{R}^d$, it computes which simplex σ of the triangulation contains q with a constant-time bucket-search point location algorithm based on FORTRAN code by Edahiro, Kokubo, and Asano (1984).[108] It then computes the barycentric coordinates of q with respect to the vertices of σ. The image p of q is then the point in the simplex in the triangulation in $\mathbf{R}^e$ corresponding to σ, with the same barycentric coordinates as those of q (see figure 17 again). The conversions between Euclidean and barycentric coordinates require negligible computation, involving merely the solving of a system of linear equations.

The special case where q lies outside the convex hull of the q_I is handled by partitioning $\mathbf{R}^d$ with a set of *ray-simplices*. Instead of one of the already described simplices we choose the unique ray-simplex which contains q, and then compute barycentric coordinates as usual with respect to that ray-simplex (figure 18). First we place a central point C in $\mathbf{R}^d$, at the centroid of the smallest axially-aligned

[108] The Walking Triangle Algorithm also fairly quickly finds which triangle of a mesh contains a query point, though it takes more than constant time (Lawson 1977). It has been generalized to tetrahedral meshes (Sambridge, Braun, and McQueen 1995).

hyperrectangle containing the hull (the hull's bounding box, in computer graphics terminology). Consider a face of the hull, *i.e.*, a (d−1)-simplex on the boundary of the hull. Together with C, its vertices define a d-simplex. The set of all such simplices partitions the hull; but if we extend each simplex away from C (imagine rays extending from C through all the vertices of the hull), the resultant ray-simplices in fact partition all of $\mathbf{R}^d$ (the faces of the ray-simplices overlap, of course, but we can ignore this in our application as their barycentric coordinates will coincide).[109] Formally, we define the ray-simplex $\rho(\sigma, v)$ of a simplex σ and a vertex v of σ as the set of points whose barycentric coordinates with respect to σ are all nonnegative, with the possible exception of the coordinate corresponding to v. There are many ways to choose the point in $\mathbf{R}^e$ corresponding to C and thereby induce corresponding ray-simplices which partition $\mathbf{R}^e$; the particular way does not affect the continuity and piecewise linearity of this extension to the mapping. We conveniently define the point in $\mathbf{R}^e$ corresponding to C, one dimension at a time: its I^{th} coordinate is the median of the I^{th} coordinates of each point on the boundary of the hull in $\mathbf{R}^e$.

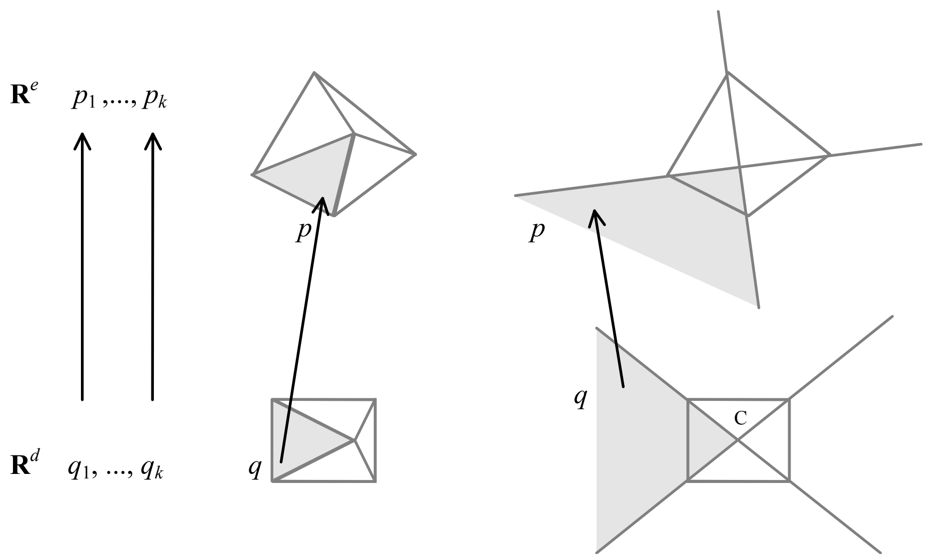


Figure 18. The centroid C and the four edges of the hull of the q_I induce four ray-simplices, one of which is shaded. If q lies outside the hull, the mapping from q in $\mathbf{R}^d$ to p in $\mathbf{R}^e$ is defined in terms of corresponding ray-simplices instead of corresponding simplices.

We now analyze the complexity of the algorithm. Initialization consists of running the genetic algorithm or Sammon's mapping (both use $O(k)$ time and space), computing a Delaunay triangulation (also $O(k)$), and initializing the bucket-search algorithm. The algorithm described by Edahiro, Kokubo, and Asano (1984) works only when $d = 2$, and takes $O(k)$ time and space. It can be weakly generalized to higher dimensions. At initialization, a regular d-lattice is filled with k^2 points; for each point we also

[109] Proof: for any point q, draw the ray from C through q. This ray will meet exactly one face of the hull, because the hull is convex. The ray-simplex induced from that face is the unique ray-simplex containing q.

record which simplex contains it. During the running phase, to determine which simplex contains a query point, we first try the simplices associated with the two lattice points nearest the query point. Most of the time these will succeed. If they fail, we use direction vectors to narrow down which simplex to choose, in the manner of the Walking Triangle Algorithm (Sambridge, Braun, and McQueen 1995). In the worst case, of course, we can scan all the simplices until one is found. The space complexity of this generalization is $O(k^2)$. Initialization performs k^2 brute-force scans. Each scan has complexity proportional to the expected number of simplices in the complex. By a theorem of Dwyer, this is $O(k)$ if the points q_I are uniformly distributed through the unit d-ball (here, a d-rectangle). A weaker condition holds, in fact: the distribution of points can be only quasi-uniform, *i.e.*, with density bounded above and below by two positive constants everywhere in the d-ball (Dwyer 1991).[110] The complexity of each scan also depends on the test to see if a point is in a given simplex. This involves a conversion from barycentric coordinates to Euclidean, a system of d equations in d unknowns, solvable in $O(d^3)$ time by Gaussian elimination. So the initialization time of the generalized Edahiro's algorithm is $O(k^2)O(k)O(d^3) = O(k^3d^3)$. To summarize: for $d = 2$, initialization takes $O(k)$ time and space; for $d > 2$, it takes $O(k^3d^3)$ time and $O(k^2)$ space.

During the running phase, the expected time to find which simplex contains the query point is $O(1)$. Computing the query point's barycentric coordinates with respect to this simplex takes $O(d^3)$ time using Gaussian elimination, as above. Converting these barycentric coordinates of a d-simplex in $\mathbf{R}^e$ to Euclidean coordinates takes $O((d+1)e) = O(de)$ time. The total time complexity for mapping a point from $\mathbf{R}^d$ to $\mathbf{R}^e$ during the running phase is thus $O(d^3 + de)$. Note that this is well behaved as e becomes large, and does not increase at all as generating points are added (as k increases).

Since the domain of the simplicial interpolator is the unbounded totality of $\mathbf{R}^d$, its range may also be unbounded. This may not be compatible with the inputs of a synthesis algorithm, where bounds such as nonnegative amplitudes and stable filter coefficients are common. Three methods can ensure bounded output from a simplicial interpolator (or any generator of unbounded data, for that matter). (I) If a real-valued component of $\mathbf{R}^e$ must remain within the closed interval $[a,b]$, it can simply be clamped: $y = x$ if $a \le x \le b$, $y = a$ if $x < a$, $y = b$ if $x > b$. (ii) If the endpoints must be avoided, a sigmoid-shaped function like an arctangent monotonically maps $\mathbf{R}$ to (a,b). (iii) In some cases valid input cannot be broken down into individual dimensions of $\mathbf{R}^e$, because of inter-dimensional constraints (stable filter coefficients, for example). If the valid subset V of $\mathbf{R}^e$ can still be computed *a priori*, a maximal convex subset of V (ideally V itself) can be dilated to cover an e-rectangle. This reduces the problem to one solvable by (I) or (ii): the output of the simplicial interpolator is some point p in $\mathbf{R}^e$. Applying a clamp or sigmoid to each

[110] An example of a distribution which fails this criterion is one whose support is a curve or sheet.

coordinate of p moves p into the e-rectangle. Inverting the dilation then carries p into V, thereby making p a valid input for the synthesizer.

3.2.2.3 Bowler's interpolator

Bowler et al. (1990) present a way to define mappings from $\mathbf{R}^d$ to $\mathbf{R}^e$ which are continuous and piecewise differentiable (in fact piecewise linear, it can be deduced). The technique takes as input a lattice in $\mathbf{R}^d$, analogous to the q_I of simplicial interpolation. At each point of this lattice is stored the image of the map (the p_I) in $\mathbf{R}^e$. This defines a pointwise map from $\mathbf{R}^d$ to $\mathbf{R}^e$. They imply that conventional extensions from this pointwise map to a continuous map (on the hull of the d-lattice) work by first finding which cell of the lattice contains a given point q in $\mathbf{R}^d$, and then constructing the image p of q by interpolating among the $\mathbf{R}^e$-values associated with the vertices of the cell. The number of such vertices is 2^d. At the expense of some additional geometrical calculations, they reduce the number of vertices among which to interpolate from 2^d to $d+1$, by dividing each lattice cell canonically into d-simplices.[111]

Let a vertex of the d-lattice have coordinates or indices $(x_1, \ldots, x_d)$ for some integers x_i. Bowler's algorithm begins by assigning an integer C_V to each vertex V of the lattice, namely how many indices of V are odd. Each C_V is in $\{0, \ldots, d\}$ because V has d indices, being in $\mathbf{R}^d$. After this initialization, interpolating (computing the image p in $\mathbf{R}^e$ of q in $\mathbf{R}^d$) is done by finding the vertex V with $C_V = j$ that is closest to q, for each j from 0 to $d-1$. They do not specify a search algorithm, but a constant-time algorithm is presented below. This search yields $d+1$ vertices V_j, which in fact define a simplex in the simplicial complex induced by the lattice.

The algorithm for finding which lattice cell contains the probe point is not given for general lattices. For regularly spaced lattices aligned with the principal axes of $\mathbf{R}^d$, the obvious algorithm has cost $O(1)$; we can take $O(1)$ as a trivial lower bound for the general case without changing our conclusions, as it turns out.

The paper does not specify the computational cost for computing which simplex within the lattice cell contains the probe point q, but an $O(1)$ lower bound is given by an algorithm proposed by Edelsbrunner for d-lattices constrained to be rectilinear and square. Given a point $q = (q_1, \ldots, q_d)$, find the "lower left" corner q' of the cube by truncating: $q' = (\lfloor q_1 \rfloor, \lfloor q_2 \rfloor, \ldots, \lfloor q_d \rfloor)$. Label each vertex of this cube with the number of cube-edges in a shortest-length path from q' to that point. (So q' has label 0, its neighbors have label 1, its neighbors' neighbors have label 2, and its antipode has label d.) As an example, take the case

[111] A note to readers of (Bowler et al. 1990): to clarify its mathematical logic, omit all phrases of the form "modulo (N+1)."

where $d = 3$ (q is in $\mathbf{R}^3$) and q' is (0,0,0). Call the three points with label 1 (along the three principal axes) p_1, p_2, and p_3. Now we find which of these p_I is closest to q'. Compare p_1 to p_2 by checking the sign of $q_{x_1} - q_{x_2}$. If this is negative, let $I = 2$ (otherwise $I = 1$); this indicates which of p_1 and p_2 is closer to q'. Compare p_I to p_3 by checking the sign of $q_{x_i} - q_{x_3}$. Again, take the closer of these two: this will be the point closest to q'. So choosing the closest point with label 1 is a 3-dimensional problem taking 2 comparisons. Given this 1-label point, we find the closest 2-label point to q' by noting that it must be one of the two 2-points adjacent to the closest 1-point. So finding the closest 2-point is a 2-dimensional problem requiring one comparison. Finding the closest 3-point is trivial since there is only one of these.

This algorithm generalizes to $\mathbf{R}^d$, can be implemented recursively for each successive k-point, and takes $(d-1) + (d-2) + \ldots + 1 = O(d^2)$ comparisons. Correctness comes from the fact that a hypercube is partitioned by the simplices induced by each shortest path from one corner to its antipode. So the "search time" for Bowler's interpolator is then $O(d^2 + d^3 + de) = O(d^3 + de)$.

An error occurs at equation (2) of (Bowler et al. 1990): V_I should read $(V_I - V_0)$. Let us correctly restate this equation and the following two (their N, V_I, P, y_P correspond to our k, q_I, q, p respectively):

(2): $P = V_0 + \sum_{i=1}^{N} \alpha_i (V_i - V_0)$

(3): N linear equations of the form $x_{P_j} = x_{0_j} + \sum_{i=1}^{N} \alpha_i (x_{i_j} - x_{0_j})$

(4): $y_P = y_0 + \sum_{i=1}^{N} \alpha_i (y_i - y_0)$

Define $\beta_0 = 1 - \sum_{i=1}^{N} \alpha_i$, and $\beta_i = \alpha_i$ for I from 1 to N. It then follows that $(\beta_0, \ldots, \beta_N)$ are the barycentric coordinates of P with respect to the $\{V_i\}$, so (4) can be rewritten as

(4b) $y_P = \sum_{i=0}^{N} \beta_i y_i$

So Bowler's interpolator turns out to be a special case of the simplicial interpolator, with generating points arranged in a particular configuration (a d-lattice), and using a fixed (not necessarily Delaunay) triangulation.

Restated in the terminology of simplicial interpolation, the system of equations (3) takes $O(d^3)$ time to solve using Gaussian elimination. Equation (4b) is directly computed in $O(de)$ time. So a lower bound for computing the image of q using this scheme is $O(1) + O(d^3) + O(de) = O(d^3 + de)$, the same as the exact cost of computing the image of q using simplicial interpolation. So Bowler's interpolator takes at least as much time as the simplicial interpolator. It differs from the simplicial interpolator in its initialization. The simplicial interpolator starts with an arbitrary set of points in $\mathbf{R}^e$, while Bowler constrains the control space to be a *d*-rectangle with internal lattice structure. The simplicial interpolator can run with far fewer generating points than the lattice of generating points needed by Bowler. If a particular application already has this *d*-lattice constraint and can afford its high $O(2^d)$ memory usage, Bowler's simpler architecture may be indicated. Otherwise, the simplicial interpolator is preferred for its greater generality and lower memory use (at the expense of greater initialization time).

3.2.2.4 Bilinear interpolator

Bilinear interpolation interpolates between four generating points *a*, *b*, *c*, and *d* in $\mathbf{R}^e$ by moving a cursor in a square. Watson (1992, 139) gives some historical background; he also calls it rectangular hyperboloid interpolation and interpolation with bilinear patches (when generating points are the corners of successive rectangles of a grid).[112] Here is an outline of the algorithm. For each dimension *k* from 1 to *e*, do the following. Take the values of the k^{th} coordinate of each point, a_k, b_k, c_k, and d_k. Make them the heights of four points directly above the corners of a unit square, $(0, 0, a_k)$, $(0, 1, b_k)$, $(1, 0, c_k)$, $(1, 1, d_k)$. To interpolate between these four points, use two scalars I_x and I_y. Draw a line *L* from *a* to *b*. Now sweep *L*, moving its endpoints gradually to *c* and *d* respectively (this means *ix* sweeps from 0 to 1). This defines a surface connecting the four points. As I_y sweeps from 0 to 1, this corresponds to a point moving along the line *L* from the *a*-*c* end to the *b*-*d* end. All *e* surfaces taken together define a two-parameter interpolation between the four original *e*-dimensional points.[113] Here is pseudocode that computes this interpolation for one of the *e* surfaces (Yu 1996):

```
Procedure InterpolateSquare(ix, iy, a, b, c, d)
   d0 = b - a
   d1 = d - c
   range = d0 + iy * (d1 - d0)
   start = a + ix * (c - a)
   return start + ix * range
```

Equivalently,

```
Procedure InterpolateSquare(ix, iy, a, b, c, d)
   return a + ix  * ((c-a) + (1-iy)*(b-a) + iy*(d-c))
```

[112] In the context of controlling parameters of sound synthesis, Maldonado (1999) calls such grids (in one, two or three dimensions) "hyper-vectorial synthesis."

Here two parameters I_x and I_y, as they range over [0,1], control the interpolation between four values a, b, c, d. This method generalizes from two to higher dimensions, to produce an interpolator where n parameters control 2^n values, naturally enough called trilinear or multilinear interpolation. This generalization requires that the number of generating points (the points a, b, c, d) is a power of two. Avoiding this requirement is difficult. For example, say the number of points is 9. We can round 9 up to 16 and introduce 7 dummy points. Neither the values of these dummy points nor their positions in the complete 16-tuplet can be chosen without introducing a high level of arbitrariness.

Bilinear interpolation might also be generalized by scattering other generating points inside the square, instead of extending the square to an n-cube. This avoids adding more control parameters to the original two. But now we must choose where in the square to place generating points beyond the first four. For that matter, which points shall be given the privilege of being the four corners? We could use a genetic algorithm to choose positions to maximize some measure of goodness like approaching the configuration of the points in $\mathbf{R}^e$. Even so, however we choose these points, we are still using only two parameters for the square. Staying in the rectilinear idiom, we induce a lattice from the interior points to divide up the square into an (unevenly spaced) checkerboard; each cell of the checkerboard is then amenable to standard bilinear interpolation. But this reduces to the interpolation scheme in (Choi, Bargar, and Goudeseune 1995), of which the simplicial interpolator is an improvement. So this breaks no new ground.

Bilinear interpolation differs from simplicial interpolation in that the number of control parameters increases as $\log_2$ of the number of generating points instead of remaining constant. We conclude that bilinear interpolation works best when the number of generating points is 4 or 8, its simplicity in these cases possibly outweighing the generality of simplicial interpolation. Of course this precludes easy local modification of the mapping by adding a few generating points here and there.

3.2.2.5 Natural neighbor interpolation

Consider a finite set of points $\{x_I\}$ in the plane. If a circle can be drawn through two of these points which surrounds no other point from the set, then the two points are called *natural neighbors*. Equivalently, two points are natural neighbors if their Voronoi cells share an edge. This definition of natural neighbors extends straightforwardly to higher dimensions.

---

[113] Such swept-out surfaces are called *ruled*; this surface (one of the e) is a rectangular hyperboloid. Note that different surfaces and different interpolations usually result if we choose a different permutation of the four points. This asymmetry is unfortunate.

For any interpolator at all, the *interpolation subset* of a point x is defined as those x_I whose values $f(x_I)$ are used to interpolate among in determining $f(x)$; in other words, $f(x)$ is computed as a weighted sum of the $f(x_I)$. In proximal interpolation the interpolation subset is the single point nearest x. A generalization of this, *k-nearest neighbor interpolation*, uses the k nearest points. *Natural neighbor interpolation* uses as interpolation subset the points whose Voronoi cells are adjacent to (or are!) the cell containing x; how many points this is varies with the density and distribution of points. To be precise, natural neighbor interpolation actually adds the point x to the set of x_I's, recomputes the Voronoi diagram, and remembers x's Voronoi cell C. Returning to the original Voronoi diagram, it considers what fraction of C is covered by each cell. The fraction corresponding to each cell (and that cell's x_I) becomes the weight for that $f(x_I)$ in the weighted sum which computes $f(x)$. (This procedure has given natural neighbor interpolation the nickname of area-stealing interpolation.) Constructing f in this baroque but robust way makes it C^∞ everywhere except at the x_I themselves, and continuous everywhere. Two seminal papers describe natural neighbor interpolation (Sibson 1980, 1982); (Gold 1989, 1990) are more recent discussions. It is perhaps the most robust interpolator possible for irregularly spaced data, but is difficult to implement in high dimensions. Simplicial interpolation can be thought of as a piecewise linear approximation to natural neighbors interpolation, one which works well in high dimensions.

From a comparison of several dozen interpolation methods in (Watson 1992, 159), it is again evident that natural neighbors is the most robust interpolator with the fewest assumptions about its input data. Watson (1992, 157) considers ways to make the interpolation function smooth even at the x_I, by "blending" the variation in gradient normally concentrated at such a point throughout a region surrounding it. Although blended natural neighbors interpolation has many advantages, it is challenging to implement; no claim of a high-dimensional implementation has been published. Simplicial interpolation I have found adequate for a wide variety of mappings with the eviolin, despite the nondifferentiable edges it produces.

3.2.3 Automated evaluation of a synthesis algorithm

The pointwise mapping from control values to sound synthesis parameters which is used by an interpolator can be computed automatically by simulating how a person tries out different control values and listens to the corresponding sounds. Once computed, this mapping can of course be manually altered afterwards.

The family of sounds produced by a synthesis algorithm can be thought of as a space defined by a product of intervals, each interval being the range of one continuous input from some minimum to some maximum. We call this space a hyperrectangle or e-rectangle, a subset of $\mathbf{R}^e$. The simplicial interpolator takes as input a small set of generating points in this e-rectangle, chosen to correspond to interesting sounds. We can go one step farther: assuming that variety is desirable for the family of sounds, optimal generating points can be computed with a technique of automated perception which I call a *timbre rover* (figure 19).

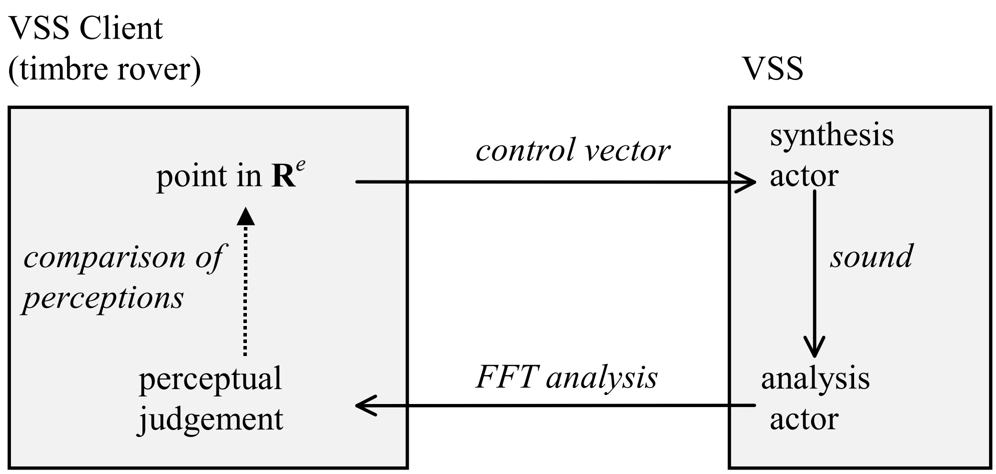


Figure 19. Data flow in the timbre rover. The timbre rover finds a set of points in $\mathbf{R}^e$ whose corresponding set of timbres has maximal variety.

A synthesis algorithm runs in VSS, a real-time sound synthesis framework developed at the National Center for Supercomputing Applications (Bargar et al. 1994). This algorithm can be any actor or collection of actors in VSS which produces an acoustic signal and is controlled by a vector of floating-point values. Control in the form of such vectors comes from a VSS client, a separate program communicating with VSS. The client sends a vector to the synthesis algorithm, causing a change in the sound. The sound is fed to an analysis actor which performs an 8192-sample FFT on it. This analysis, the output of the FFT, is sent back to the timbre rover. So the client repeatedly tries out different vectors (points in a hyperrectangle) and listens to their resulting sounds. This analysis is certainly incomplete; in particular

FFT analysis discards time-variant information like attack transients. The analysis actor actually waits a short while for transients to dampen after a new control vector is sent, before beginning the FFT.[114]

The timbre rover takes as input a synthesis algorithm, a minimum and maximum value for each of its *e* continuous (real-valued) input variables, and a small integer *d*. It requires no other input, in particular no knowledge of the synthesis algorithm or any relationships between parameters.[115] Via simplicial interpolation it produces a continuous map from a subset of the *d*-rectangle to the *e*-rectangle, continuous in the classical sense of a map from $\mathbf{R}^d$ to $\mathbf{R}^e$, but also continuous and linear, as nearly as possible, considered as a map from $\mathbf{R}^d$ to the perceptual space P. (Continuity is here defined in terms of perceptual distance in P.) Continuity means that nearby points in the *d*-rectangle correspond to similar timbres. Linearity means that variation of timbre is spread evenly throughout the *d*-rectangle rather than being concentrated in small regions; this uniform distribution of timbre corresponds to maximum entropy and hence maximum (expressive) information.

The sounds from the *e*-rectangle are assumed to be reasonably behaved, perceptually smooth and preferably perceptually convex (if two sounds can be produced, so can any sound "in between"). If a point moving smoothly in a rectangle produces sounds which vary wildly or have many discontinuities, no scheme can successfully simplify or even search that rectangle.

The timbre rover works best for values of *e* between 4 and about 15. It is overkill for fewer dimensions, where conventional adaptive mesh refinement (AMR) applies.[116] AMR fails in more than 3 dimensions, because in such spaces refining algorithms have been found to converge too slowly or too inaccurately. For more than 15 dimensions, extra assumptions about the structure of the space are required to deal with its sheer size.

The timbre rover chooses which points in the *e*-rectangle to try out by trying to find a set of *k* generating points whose corresponding timbres differ maximally. Obvious first points to choose are the corners of the *e*-rectangle. But even with this restricted set brute force fails: 20 synthesis parameters produce 2^{20} or about 10^6 corner points. Some method is needed to prune unpromising regions from the search and

---

[114] Temporal variation of sound can be identified with composition, with musical form on smaller and larger scales. A timbre rover that compares time-varying sounds is not too far from one that compares actual compositions!

[115] To quickly explore the perceptual space, there is one property which the synthesis algorithm should possess: changes in its input parameters should affect its acoustic output quickly, certainly within 300 msec if any change happens at all.

[116] AMR is described in (Berger and Oliger 1984; Berger and Colella 1989; Bell 1994). An excellent online introduction is found in (Mitra, Parashar, and Browne 1998).

focus on areas with greater variety.[117] An octree-style decomposition of the hyperrectangle is possible: if the hyperrectangle were the unit k-cube, choosing points with coordinates only 0 or 1 except for one being ½, then two being ½, and so on (edge centers, face centers, *etc.*), then allowing coordinate values of ¼ and so on as the octree cells refine. But the number of cells grows even faster than the number of corners: even only 10 parameters down three levels (9 steps of resolution along each dimension) lead to $9^{10} = 3.5\times10^9$ points. As the analysis of sound for each point takes significant real time (about 300 msec), aggressive pruning is required to finish a run of the timbre rover in days instead of decades.

Instead, the timbre rover starts with a set of k points and then iteratively improves it. We say that a set X is better than (improved from) another set Y if the points of X are farther apart, which means that X's vector of $\binom{k}{2}$ distances, sorted in increasing order, is lexicographically greater than Y's. The set of points is improved by discarding the point closest to the others and replacing it with a point found to be farther away than the discarded one was.[118] (We analyze the points chosen only as needed, and cache the resulting analyses in a hash table. Hashing the computed differences between points is ineffective because it uses so much space: even simple examples use hundreds of megabytes of memory, and the avoided recomputation is more than offset by increased page swapping.) This whole process is done several times, and the best overall set produced is kept.

We now consider two subproblems: (i) how to define the difference between two sounds, and (ii) where to look for a promising new point to improve the set of k points.

Each point in $\mathbf{R}^e$ (*i.e.*, sound) has some measurement describing its percept. Since we discard temporal information and consider only its steady-state behavior, this measurement is its frequency content, its Fourier transform. The synthesizer can be said to map vectors of synthesis parameters, points in $\mathbf{R}^e$, to perceptual points in a space P; in other words, the synthesizer is a "black box" transfer function from $\mathbf{R}^e$ to P. (We assume only that P is a metric space. Beyond this, P may be isomorphic to a subset of some Euclidean space of sufficiently large dimensionality, but this is only incidental. We never explicitly represent such a Euclidean space in what follows.)

The timbre rover compares the analyses (complicated numerical structures) of two sounds to form a scalar measure of perceptual distance. Experimentally, this distance function appears to be a metric. It is easy to verify directly from the software that the distance between a point and itself is zero and that the

---

[117] If the timbre rover has any aesthetic assumption, this is it: for any given purpose, most of a high-dimensional space is uninteresting. The search which the timbre rover undertakes is one of avoiding the boring wastelands.

function is symmetric; the triangle inequality appears to hold, though it is not clear how this could be proved for all possible synthesis algorithms. This distance function is based on existing work in psychoacoustics. Feiten and Günzel (1993) describe the training of neural networks called Kohonen Feature Maps to classify sounds based on perceptual difference, starting with experiments on human listeners. The sounds they used were steady-state harmonic spectra based on 300 Hz, using only the first ten partials. Amplitudes of individual partials were constrained to approximate spectral contours of actual orchestral instruments. The formal description of a sound was a vector of the loudnesses (in sones) of the ten partials. They found that the distance function most nearly approximating their listeners' perception was the Minkowski metric with $p = 5$.[119]

As this model breaks down for greater numbers of partials when several partials occupy the same critical band, the timbre rover modifies this model to use critical bands. The decibel amplitudes reported by the FFT are adjusted by the Fletcher-Munson curve for 80 dB above the threshold of hearing, to approximate a value in sones.[120] This produces a loudness-versus-frequency plot of the sound. The timbre rover then divides this plot into critical bands and notes the loudness present in each band. Each band is separated into an overall noise floor and zero or more spectral peaks, where noise floor is defined as the median loudness of all frequencies in the band. Spectral peaks are defined as local maxima exceeding the noise floor by more than one average deviation; each peak consists of a frequency, loudness, and width. Since the first six partials of the sounds studied by Feiten and Günzel occupy entirely separate critical bands, the Minkowski metric can reasonably be extended from vectors of loudness of individual partials to vectors of loudness of critical bands. The timbre rover thus computes the distance between two sounds by computing the Minkowski distance between their vectors of critical band loudnesses.

This perceptual model disregards (does not explicitly represent) pitch. If all sounds presented have the same pitch, this is moot; various sets of generating points can be defined for different pitch areas if needed, as with the various registers of the clarinet. On the other hand, this model does not disregard amplitude. Many writers consider musical tones to consist primarily of three independent attributes, pitch, amplitude, and timbre. For our purposes, we need not separate from the overall perceptual

---

[118] Discarding the point closest to the others is like discarding a weak card in a poker hand—though if the timbre rover played poker it would search for a hand of maximally different cards and therefore try hard to lose.

[119] The Minkowski metric of two vectors x and y in $\mathbf{R}^k$ is defined by $d_M(x, y) = \left(\sum_{i=1}^{k} |x_i - y_i|^p\right)^{1/p}$. The city-block metric and Euclidean metric are special cases, corresponding to $p = 1$ and $p = 2$ respectively.

difference between two sounds the particular perceptual difference attributable to difference in overall amplitude. Normalizing amplitudes (to peak amplitude? to a common perceptual loudness?) adds arbitrariness to the model without contributing strength.

But it seems presumptuous to apply the mathematical ideal of a scalar difference measure to something as messy as sounds in general. More realistic may be a vector difference derived from various scalar differences (critical-band loudnesses as above, absolute pitch interval, pitch class interval, certain properties of single critical bands like roughness due to multiple partials present, time-variant behavior), *not* subsequently reduced to a single scalar. Langmead (1995) evaluated nine separate measurements of timbral distance and found it very difficult to meaningfully combine these into a scalar distance.

Individual elements of this difference vector may vary in prominence depending on the musical context. If an element has been varied little, a sudden increase in variation (even a single change) attracts attention. Continued variation of this element does not hold the listener's attention, though. (Formulating this observation in mathematical terms would require considerable psychoacoustic experimentation.)

Let us now consider where to look for a promising new point to replace one discarded from a set of k generating points. We can simply choose a point randomly, uniformly, from the hyperrectangle. (Even if it turns out to not be useful here, its analysis may be used in later passes of the algorithm starting with a different random set of k points.) This is in fact what the software does at present. For some sound families it works quite well, gradually improving the set. For others it fares poorly: after a few steps, choosing even a hundred points fails to find any better points. This may be because the space has large regions of approximately equal sounds. In such cases it may be better to move the discarded point slightly instead of replacing it with a distant point.

How many generating points should we use? How large should k be? A very large value, say 10^4, will adequately fill the space of possible sounds. But we cannot assume any correspondence between the shapes of the space of sounds and the controlling space, a d-rectangle. The resulting map between the spaces, defined by the simplicial interpolator, may be highly corrugated (having regions of high curvature or even nondifferentiability) and therefore be difficult to perform with. A neural gas algorithm or something similar would be needed to discover the actual topology of the sound space and thereby construct a similar control space (Martinetz and Schulten 1993). On the other hand, the minimal value of $k = d + 1$ may impoverish the variety of sounds. So I propose a middle route. One can begin with this

---

[120] This curve is in the middle of the ear's amplitude range and changes the amplitude values rather mildly. In no case (for no amplitude, loud or soft) is it worse than the original flat frequency response. This psychoacoustic data is taken from (Dodge and Jerse 1984, 45).

minimal value to find an initial set of points. This now-fixed set of points can be extended with more points, perhaps twice as many, chosen from the ball bounding the *d*-simplex defined by the fixed points. This extension procedure can be repeated many times, choosing points from the balls bounding each simplex in the complex defined by Delaunay triangulation, until a sufficient number of generating points have been found to result in a good map. Restricting the region from which new points are chosen prevents the map from becoming excessively corrugated. At every step of the map's refinement, it looks approximately like a *d*-simplex, never approaching the high curvature of something pathological like a space-filling curve.

3.2.4 Effectively reducing dimensionality

Stated colloquially, reducing the number of dimensions of control makes an instrument less scary for the performer. This is a powerful contribution which high-dimensional interpolators can make to a synthetic instrument. Here we consider how any HDI can best be used to this end.

3.2.4.1 Cross-coupled controls

A set of controls is *cross-coupled* when it drives a set of dimensions in such a way that the individual controls cannot be separated. The traditional example of this in human factors literature is piloting an aircraft: throttle and elevators both drive airspeed and rate of climb; ailerons drive yaw, roll, and rate of climb; rudder drives yaw. Obtaining a particular heading, airspeed, and rate of climb requires the pilot's combined use of all these controls.

Using the terminology of the previous chapter, cross-coupled controls are those which lie in the same connected component of the driving graph (figure 10). The greater the connectivity of that component, the more strongly cross-coupled those controls are. The several muscular controls a violinist's left arm has over pitch are connected through only the single dimension of pitch, so this component of the graph is only 1-connected. By contrast, vowel color in singing exemplifies a multiply-connected graph component: arguably all 22 muscles controlling mouth shape drive each of the five strongest formants of a vocal spectrum, producing a 5-connected component. The mechanics of a violinist's right arm are also strongly cross-coupled. We can call shoulder rotation, elbow bend, wrist bend, and axial forearm rotation four controls; these drive four dimensions, longitudinal bow position, distance from bow-string contact point to bridge, angle between bow and string, and downward force of the bow on the string. There are still other dimensions (which string is bowed, axial bow angle) but the point is clear: each control drives several dimensions, and each dimension is driven by several controls.

But there is a second level. After the intra-arm controls drive this set of physical dimensions (angle between bow and string, *etc*.), the physical dimensions in turn (now considered as controls, to abuse our terminology) drive a cross-coupled set of acoustic dimensions. We begin to see why bowing takes such a long time to master and why few violin-playing robots have been made.[121] In determining how the bowed string vibrates, we cannot consider in isolation bow speed, bow force and distance from bridge. These three parameters define a space which contains regions of normal tone, sul ponticello, flautando, and so on. None of these regions can be reduced to a region of fewer than three dimensions projected along the other axes. (Schelleng (1973) provides diagrams of these regions.)

If we consider the relative influence of each control on a given dimension, we often see that one control drives the dimension more strongly than the others, and that that dimension is driven most strongly by that control. For example, the throttle of an airplane is most strongly coupled to airspeed, and the bend of a violinist's right elbow is most strongly coupled to longitudinal bow position. Mathematically, we can construct a matrix of coefficients indicating the influence of each (control, dimension) pair. A matrix with greater diagonal elements and smaller off-diagonal elements (assuming a permutation of rows and columns which maximizes this) indicates a stronger specialization of controls to dimensions and a correspondingly weaker cross-coupling.

Cross-coupled controls are more difficult for performers to learn because part-task training on individual controls transfers poorly to the whole task. But experiments by Hunt and Kirk (1999) indicate that cross-coupling can produce a better controller, once learned. If a new instrument design seems to demand cross-coupled controls, grouping them together as inputs to an HDI encourages the performer to think of the controls as a single entity. The converse holds, too: a set of strongly non-coupled controls like a large bank of sliders is inappropriate for input to an HDI. This is because the input devices misleadingly suggest that each one has an inherent meaning.

3.2.4.2 Networks of controllers

The outputs of controllers can, instead of driving acoustic dimensions directly, be connected to the inputs of other controllers. In this way networks of controllers can be assembled by connecting them in series or in parallel.

Sometimes it is appropriate to reduce the number of dimensions from e to d in several stages instead of all at once—a set of *cascaded* controllers. This is particularly appropriate if a strong grouping of

---

[121] Shibuya et al. (1997) describe research directed at this goal. Another violin-playing robot simplifies the mechanics of the bow arm: it bows individual strings by rotating the violin about its longitudinal axis (Kajitani 1992).

parameters exists in $\mathbf{R}^e$, in which case several controllers, one per grouping, can be used in parallel and then cascaded to another one. For example, consider a synthesizer with three components each having many dimensions. Each component might be individually reduced to two dimensions, with the resulting six dimensions then being reduced to two final dimensions.

Using controllers in parallel is suggested by orchestral instruments which have to a large part independent controls for separate acoustic parameters, typically pitch and timbre as with woodwinds and strings. As pitch is so primary in human acoustic perception (fineness of resolution), synthetic instruments may similarly find it advantageous to decouple it from other sonic parameters in the interface presented to the performer.[122] Indeed, the *ondes martenot* and later the Minimoog synthesizer used one hand to control pitch on a conventional keyboard and the other to control aspects of timbre (or pitch bend). This separation of control of pitch and timbre between a bank of keys and a high-dimensional control is found in two other designs as well. The Electronic Sackbut analog synthesizer controls timbre with a simultaneously selectable bank of pressure sensors and touchpad (Young 1989, 175–179); the aXi0 general-purpose MIDI controller controls pitch with a woodwind-like chording keyboard, and timbre with a multislider, a high-dimensional joystick (Cariou 1994). Coarse separation of pitch and timbre is effective enough to have made so-called mod wheels a standard part of MIDI keyboard controllers.

High-dimensional interpolators may be cascaded to better reflect the cross-coupling structure of synthesis parameters. They can also be run in reverse as dimension-reducing mechanisms, providing tractable ways to drive one high-dimensional system with another, say a 20-to-3 interpolator driving a 3-to-10 interpolator. Of course this loses more information than a straight 20-to-10 interpolator; but in such a situation information may be so abundant that some loss is necessary and desirable.

The value of any dimension-reducing controller is found exactly in how well it loses information. A controller which tries to preserve all information (such as a complete set of linear sliders) will paradoxically still lose information because it is difficult to use. A synthesizer with many degrees of freedom can be played in an *ad hoc* manner by a finitely attentive human performer, exploring first this and then that region of its parameter space. But richer musical results are more probable if, with regard to information, the synthesizer's rate of consumption is systematically matched to its performer's rate of production. Stated another way, the performer can do only so much. Controlled loss of information is about deciding what the performer can and cannot do, about matching that dividing line with the one between expressive and inexpressive.

---

[122] Certainly orchestral performers desire this: better woodwind instruments play in tune with less embouchure correction.

3.3 *Physical and psychological issues*

3.3.1 Human factors design

We now consider the elementary capabilities and limitations of the human performer (the "human operator" executing a task) in terms of control theory, a formal description of what we just alluded to, what the performer can and cannot do.

Performers have a limited rate of transmission of information when responding to input and producing output (respectively, evaluating the sound of an instrument and playing it). Based on this limited rate, Sheridan and Ferrell (1981, 96) suggest some guidelines for interface design: performance is generally insensitive to the input encoding used; some parallelism is possible in executing tasks; when this information channel is saturated, only a trade-off between speed and accuracy is possible.

Generally, along one dimension a person can reliably recognize about 5 to 9 levels ($\log_2(8) = 3$, so about 3 bits of information); more for musicians with near-absolute pitch (about 6 bits), less for some other perceptions. If several dimensions are presented at once, the overall level of discrimination is greater than that of each individual dimension but not as great as simple summing would predict. But redundancy, presenting the same data in several dimensions at once, does significantly increase discrimination (Sheridan and Ferrell 1981, 97–106).

Switching attention between sense modalities, or within one sense, takes about 200 msec (Sheridan and Ferrell 1981, 93).[123] Reaction time to a stimulus has been experimentally measured to be around 400 msec; it increases linearly with the rate of information transmission from the subject's input stimulus to his output response, an extra 200 msec per bit. Reaction time is otherwise independent of the number of possible stimuli or responses, but it is greatly reduced by reflex actions and extended rehearsal (Sheridan and Ferrell 1981, 114).

Performers can increase their data transmission rate by taking advantage of how short-term memory effectively recognizes short patterns (called *chunks* in human factors literature). This means grouping inputs into larger, already familiar, structures—grouping binary digits into hexadecimal digits, notes into chords, several scalar parameter values into a vector (Sheridan and Ferrell 1981, 124).

Extrapolating from experimental results presented about typists, the performer of a musical score must either (i) memorize the layout of the instrument, (ii) memorize the score, or (iii) switch attention between the two (Sheridan and Ferrell 1981, 128). Most performers of orchestral instruments choose (i); because of the visual feedback required to perform large leaps in the Romantic repertoire, pianists chose (iii) for a while and eventually moved to (ii), performing from memory. The compass of organ and harpsichord manuals did not grow like that of the piano, so organists still choose (i). Some contemporary compositions are too intricate to be memorized but still demand looking at the keyboard, forcing pianists back to (iii). Synthetic instruments therefore do well to not rely on focused visual feedback for accurate performance.

The speed and accuracy of such large leaps with visual feedback has been codified as information transmission: the endpoint of the leap, within a given tolerance, over a certain range, defines an amount of information. For example, striking a single note uniformly chosen from the piano keyboard is $\log_2(88) \cong 6.5$ bits. (Continuous tasks like lip tension or position on the fingerboard are also measurable in terms of bits; the amount of information depends on the value's probability distribution function, commonly assuming band-limited white noise or Gaussian distributions.) Fitts's Law (Fitts 1954) states that the time to perform a leap is a constant plus a linear multiple of the information transmitted by the leap; 10 bits per second is typical.

Rate of information transmission reduces gradually or sharply when the input to the performer begins to exceed what he can process. Sheridan and Ferrell (1981, 133) list the ways an individual may cope with such a situation (my spacing):

> "[O]ne might
> (1) fail to respond to (*i.e.*, omit) some inputs,
> (2) respond less accurately,
> (3) give occasional incorrect responses,
> (4) store the inputs and respond to them as time permits,
> (5) systematically ignore (*i.e.*, filter) some features of the input,
> (6) recode the inputs in a more compact or effective form, or in extreme cases
> (7) quit.
> ...[A]t moderate rates of information input overload, all methods were used about equally. At high input rates, well beyond the individual's capacity, omission and filtering were the only important methods of coping."

All these methods are familiar to performers sight reading at an externally mandated tempo which is too fast for them. Respectively, performers may (1) drop notes, (2, 3) play wrong notes, (4) play notes late, (5) ignore dynamics or ornamentation, (6) change intricate figuration into block chords, or (7) simply stop playing. This variety of coping mechanisms shows that we cannot absolutely specify how many

[123] We hear an unusual sound, and then look in that direction; or we attend to a choir's accuracy of intonation and a moment later to its uniformity of vowel color.

independent parameters a performer can attend to and control simultaneously. Rather, there is a compromise: as the number of parameters increases, accuracy of control decreases. (Increasing the number of performers of an instrument to more than one obviously increases the number of controllable parameters. But this is also a compromise: more performers create greater communications overhead between each other. Introducing extra structure not inherent to the instrument, such as pulsed rhythm or a central authority, may also be necessary for adequate control.)

In a general context, Hunt and Kirk (1999) warn computer users faced with a control task that it is unrealistic for them to expect to master it in a few minutes. Driving a car, playing billiards or soccer or darts, are all control tasks which people understand take time to learn; but players of video games tend to be less patient.[124] This computer effect carries over to music instruments: the performer expects that the computer will do everything for him, that he can do what is familiar and the computer will adapt to him. But such adaptation will always have a limit. If, conversely, the performer does not adapt, then he would be better off with his familiar instrument for familiar gestures: to that instrument he has *already* adapted, though he may have forgotten how much work that was. The more a synthetic instrument extends the orchestral instrument it is based on (if such is the case), the more the performer needs to adapt to it and learn a new set of not yet familiar gestures. For example, Trueman and Cook (1999) note that their R-Bow can make sound by simply pressing on a string without moving; I have demonstrated Theremin-like waving in space with the eviolin's own position-tracked bow. Such gestures extend traditional bowing so radically that existing gestures do not apply at all. In the limit, if the synthetic instrument is entirely new, far more rehearsal time can be expected as Bauer and Foss (1992) found with the GAMS motion tracker. Skimping on rehearsal time with synthetic instruments is perilous, even if technical components of the instrument (such as the development laboratory or a virtual environment) are difficult to reserve for extended periods.

3.3.2 Expressiveness and variety of sound

The designer of a new instrument would like it to be *expressive*. How can expressiveness be defined? In the context of a composition, the combinatorial explosion dwarfs expressiveness of the instrument itself. One cannot make a blanket judgement that flute pieces are better than trumpet pieces, for instance. The expressiveness of an instrument is more limited: what ranges of dynamics, pitch, and timbre can be extracted from it, how independently variable these ranges are, and what resolution and temporal resolution obtain within these ranges.

Expression is commonly defined in terms of representing something: empathic or plainly exhibited portrayal, delineation, depiction, rendering, setting forth (a meaning, thought, state of things), conveying a notion (Simpson and Weiner 1989). But denotation—literal representation—is overshadowed by connotation, in nonvocal music at least. Even strongly programmatic music like certain works of Berlioz and Liszt is still rendered as music, not as text.

The "expressive" *Empfindsamkeit* of C.P.E. Bach emphasized contrast from moment to moment, reacting against the baroque *Affektenlehre* and pointing towards the Classical aesthetic of opposition (manifested primarily in tonic/dominant opposition, secondarily in variety of surface rhythm and phrase length). But literary or musical art can have too wide a range of emotion; we denigrate such art as "over the top" or "schmaltzy."

If human senses had unlimited range and unlimited resolution, the choice of range in a work of art would be arbitrary. Visual contrast levels, pitch range, and dynamic range would not matter. The engineer-turned-artist thus prefers to maximize the signal-to-noise ratio by exploiting the full range of sensitivity of human input transducers (ears and eyes). If representation, communication of some pre-existent idea, has any part in expressiveness, then increasing the accuracy of this communication is justified.

And indeed, as music technology has improved, its range of dynamics and pitch has increased from that of medieval and Renaissance vocal music through the late romantic orchestra to electronic synthesis, which finally exceeds the limits of human hearing in pitch and dynamics, both high and low. The range of timbre of instruments, on the other hand, was reduced in range from medieval to romantic times as orchestral blending became more important (for example, compare vielle, shawm, krummhorn, and serpent to violin, oboe, bassoon, and tuba respectively). But again with electronics, since those times timbral range has expanded enormously.

Expressiveness is, however, more than just a matter of raw range. In the piano sonata literature Mozart accomplishes as "much" as Liszt with fewer notes, a smaller keyboard, and less dynamic range. At the time, like many composers, Mozart used all the forces at his disposal. But today we can look back at Liszt, Mozart, and even farther, and deliberately choose a reduced range of acoustic material. Visually this is like choosing black-and-white photography or charcoal sketching even when color would be no harder technically or economically: choosing subtlety over variety. The senses are not saturated. The performance is not "flat out" with nothing left to spare. The intent listener is free to attend to matters

---

[124] This is exacerbated by the relatively short life span of software-related skills as compared to skills in more slowly evolv-

other than the acoustic surface:[125] in a Mozart piano sonata subtle rubato and chord voicing play a greater role than in a Wagner opera; an Ansel Adams photograph displays texture and spatial composition better than the equivalent color photograph (or a movie, for that matter); masters of *Ikebana* floral arranging restrict their work to shades of green.

When engineers design a bridge, they specify the maximum load it can carry. The principle here is deliberately designing what something cannot do, as well as what it can do. Similarly, a synthetic instrument should be designed to do only certain things: make only certain sounds, control only certain aspects with only limited accuracy. Otherwise its software becomes unwieldy, like the supporting concrete of an overly strong bridge. Expressiveness comes at a price.

With Mozart the senses may not be saturated, but the attention is. Either Liszt or Mozart—a large or a small range of gestures—can saturate the listener. One can compose with small or large forces, with a small or large variety of gestures, since the size of the combinatorial explosion dwarfs the size of the set of gestures.

A color movie (of equivalent contrast and resolution) *can* express anything which a black-and-white photograph can, simply by abstaining from hue and motion. The argument against doing such is principally one of economics and practicality. Visually, this discussion is more about the medium (color movie or black-and-white photograph) than the individual image; musically, this discussion is more about the instrument *per se* than about individual compositions. The expressiveness of a single musical instrument, abstracted away from individual compositions, is a potential expressiveness: what the instrument *can* do, not what it *must* do at all times. If one instrument can do everything another can and more, then it is more expressive (but only if the second instrument does not already exhaust the attention of the listener or performer).

So we have defined expressiveness in terms of acoustic capability. Capability is then defined in terms of acoustic dimensions, discussed in chapter two: continuums such as frequency, amplitude, spectral brightness; their overall range and their fineness of control; and their independence of control.

For a purely mechanical system, without human physical and attentional limitations, the analogous capabilities of a communications channel can be made as great as needed: more wires, faster transistors, better error-correcting codes. But the musical instrument presents a messier problem. It has two

---

ing technologies.

communications channels, not one: from performer's muscles to device and from device to listener's ears.[126] The first channel must deal with the complicated constraints of the performer's muscular gesture and attentional limits, the second with the complicated constraints of the listener's psychoacoustic processing (masking, variable sensitivity, *etc.*) and attention.

We conclude that in designing a musical instrument, increased expressiveness is not just a matter of increasing bandwidth, adding more and more sensors and triggers and touchpads and keyboards and hoping for the best. One must also match impedances with the performer and the listener. Since the technology for synthetic instruments exceeds what performers can do and what listeners can hear, maximizing this expressiveness is a complicated nonlinear problem. We have bits and pieces of machinery to solve the problem, but at present we can only offer heuristics. The problem remains an open one.

3.3.3 Virtual environments

Virtual environments (VE's) are similar to synthetic musical instruments in several ways. Both accept gestures as input; in both cases this input is mediated by software and results in sounds, possibly images, and possibly even more. The user has demanding expectations that a VE behave like the real world in at least some ways. These are like the properties which a synthetic instrument is expected to share with orchestral instruments: low latency, repeatability, consistency, and (piecewise) continuity (Garnett and Goudeseune 1999). As these ways also tend to be the ones which most noticeably fail in a VE, a consideration of VE's will inform the design of synthetic instruments. On the other hand, VE's and synthetic instruments are also expected to differ from familiar experience in certain ways; these are typically where the user feels that the system is not failing but confusing. How a VE can overcome such confusion and overwhelming—by training or by good design—also applies to synthetic instrument design.

3.3.3.1 Perceptual-motor coupling

The gain of a control should be chosen to match the size of the performer's gestures to the change of value in the system. Small gain demands that the performer make larger gestures, perhaps ones beyond

---

[125] Reduced variety also implies greater unity of acoustic material. Individual instruments are then easier to identify. Certain sampling-based genres of popular music like *ambient* and *rave* take this to an extreme: each instrument does no more than trigger a sampled sound. The listener is then free to ignore the music (consciously and attentively) and approach a trance-like state. Any expressive variation of the sound sample would distractingly draw attention to itself.

his reach. Large gain makes the system harder to control because the required smallness of gesture surpasses the performer’s limits of accuracy. For example, in one configuration of the eviolin amplitude drove spectral brightness with fairly high gain in the hope of providing a wide timbral range without adding extra controls. It turned out that such fine control of amplitude was quite difficult; what felt to the eviolinist like five identical bow strokes actually sounded like five markedly different timbres. Even a single smooth bow stroke had objectionable timbral wobbles. Gain had to be reduced until similar bow strokes really produced similar timbres. The reduced timbral range was then increased after all by means of another control.

Gain of feedback information (*e.g.*, field of view in a VE) similarly needs to match what the performer needs to know: too small a gain and he learns much about a small part of the system, too large and he learns little about the whole system. As an example, consider visual feedback of position information for a spatially tracked control, where crossing a threshold into another spatial region produces a change of hue or intensity of a light mounted beside the control). With too small gain the performer learns only a few centimeters from the threshold that he is near it, but he can then adjust distance to that threshold to the millimeter (which is beyond his accuracy anyways). With too large gain, he knows already a meter away that he is near it but with an accuracy of only 10 cm (and that much he knows already from seeing where he is standing on stage). An intermediate gain provides useful information, detailed beyond the performer’s own sense perception but not so detailed that a gap exists between the natural senses and the augmented senses we call feedback. Unlike orchestral instruments, synthetic instruments offer the possibility of adjusting gain, to suit individual performers or even to suit different passages of a single performance.

We have already discussed the order of a control. In fact a control may have several orders at once: a more general description of how output is affected by input is given by a constant-coefficient linear differential equation (Dean and Wellman 1991, 144; Sheridan and Ferrell 1981, 178–180). The values of the coefficients then indicate how much gain is present at each order: a control can be simultaneously proportional, integral, and derivative, with a different gain for each of these three. But for our purposes we can restrict ourselves to gains which have only one nonzero order coefficient.

High system latency is of course undesirable: even more than high gain, high latency makes a system difficult to control. Baker and Wickens (1995) report that pilot performance in flight simulators varies strongly with latency, even for latencies under 50 msec. Designers and players of 3-D video games in the

[126] Singing is the obvious exception. One might say that all mechanical instruments simply realize the wish to sing better in some way.

late 1990's quickly learned that low latency (high frame rate, in their parlance) was the most important factor in realism and good game play. Of course, aiming for latency beyond human perception is wasteful. About 10 msec seems sufficient for static displays. For motion-tracked devices such as head-mounted displays or for the display of rapidly moving objects, one half to one third of that suffices. At the other end of the scale, Airey (1990) found that VE's drawn from the user's point of view (*egocentric* as opposed to an *exocentric* "you are here" cursor-on-map view) became unusable when latency exceeded about 150 msec.

Hysteresis can usefully filter noisy input (though it complicates rigorous analysis of the system and adds latency). For example, as a brass player slowly increases lip tension, the instrument continues to resonate at its current frequency for some time. Eventually it switches to the next higher overtone, where it will continue to resonate even if lip tension is brought back to the region where the lower overtone was resonating. This inaccuracy in tracking lip tension means that no single function from lip tension to pitch can be drawn. It also means that maintaining one pitch is somewhat easier, and performing lip trills is somewhat harder. More generally, hysteresis increases the simplicity of performing certain (novice-level) gestures on an instrument, at the expense of other gestures.[127]

3.3.3.2 Guiding the user

A common cause of user confusion in a VE is one of *mode*, forgetting which of several states the environment is in. Different states mean that different commands are available to the user, often from the same controls if the number of input controls is limited; upon pushing a button the user may therefore expect one result but get another. Visual indicators of mode may be ineffective since they may interfere with the environment itself; they may also be situated outside the user's field of view at the moment they are needed. Beyond awareness of what mode the system is in, switching between modes such as navigation, object selection, object manipulation, and environmental commands carries significant cognitive overhead (Smith et al. 1998). In the context of musical instruments, therefore, the designer does well to avoid a proliferation of secondary controls not reachable during normal playing.

One traditional solution used in CAVE environments is audio feedback of user actions and system state (Brady et al. 1995). Since private audio feedback to an instrumentalist can obviously be confused with or masked by the instrument's primary output, feedback in musical instruments is best presented haptically or visually. In haptic feedback we restrict ourselves to pressure and frequency; other tactile

[127] When combined with basins of attraction for detents in a mapping from **R** to **R**, hysteresis produces a quasi-fretted control which simplifies the performance of certain gestures.

senses like temperature are too slow and coarse for musical applications. Both visual and haptic feedback provide absolute identification of steps on a “gray scale” of about 5 to 7 steps, if this kind of feedback is needed. Visual feedback is easily implemented with lights, while haptic feedback frees the musician’s eyes for other tasks. Haptic discrimination of frequency is about twice as fine as visual discrimination of brightness, and changes of frequency are perceived a little faster than changes of brightness (10 to 50 rather than 20 to 50 msec), two more points in favor of haptic feedback.[128] Haptic discrimination of absolute pressure is poor, about one tenth that of frequency. Under ideal conditions spatial resolution is better visually than haptically, though: the eye discriminates points one minute of arc apart, while touch discriminates points 2 mm apart (Sheridan and Ferrell 1981, 261–262). A small visual display could therefore present tens to hundreds of times as many points as a small fingertip display.

Experimental work in musical applications of haptic feedback has centered on force-feedback keyboards, where each key has a position sensor and a mechanical driver. This research has not yet moved from the lab to the concert hall, though. Desiderata listed for such keyboards by Gillespie (1992) include low inertia, high stiffness, low backlash and the ability to be “back-driven.” Gillespie also notes that these desiderata are difficult to combine with the simulation of hard unyielding surfaces. His keyboard is built in the large, to be played with fists rather than fingers. It senses position, velocity, and downwards force on each key. By combining this data with the known power applied to the motor driving the force feedback on a key, it computes the acceleration of the key directly without the numerical inaccuracy of integrating velocity data. A different approach is found in the physically reconfigurable keyboard of Cadoz, Lisowski, and Florens (1990). Its custom electric motors/sensors combine fast response with the power needed to simulate hard surfaces. This team found a latency of 1.25 msec to be not excessive, and indeed necessary for feedback to allow adequate finesse of performance. To produce finesse in a different kind of performance, the force feedback on the control yoke of some commercial flight simulators runs with latencies as low as 0.2 msec (Ricci 2000). Force feedback differs markedly from position feedback (the latter is found for example in the motorized faders of professional audio mixing consoles). Position feedback involves the motion of masses, whose inertia on the scale of the human hand means that a latency of 10 msec usually suffices. Force feedback may involve large forces with negligible damping from inertia, because the range of motion is so small; this lack of inherent damping drives the need for very low latency.

[128] Generally, visual feedback dominates for novices while haptic and proprioceptive feedback dominates for experts (Vertegaal, Ungvary, and Kieslinger 1996, 310). This is the common experience of music students chided by their teachers to not look at the fingerboard or keyboard or pedalboard. A synthetic instrument would ideally have both kinds of feedback, as the performer matures.

Leaving the world of keyboards, the Moose is a computer mouse offering force feedback (a planar haptic interface). Its workspace is 8 cm square with 12-bit resolution. Maximum force is about 25 N, passive inertia is a rather high 1 kg, but it runs at an impressively fast 1 kHz (Miner, Gillespie, and Caudell 1996; O'Modhrain and Gillespie 1997). Chafe and O'Modhrain (1996) describe a musical use of the Moose, displaying the "effort" of performance of a short excerpt of a Beethoven piano sonata. They conclude that force feedback can usefully reduce many parameters to a single summary parameter for general nondetailed feedback about a system. Haptic feedback in the form of vibration rather than position is given by the VR/TX system (Rovan and Hayward 2000). This system is designed expressly for non-contact motion tracking, what they call open-air controllers. Experimentation with the VR/TX system agrees with previous research: its useful frequency range is 70 to 800 Hz, within which most people can distinguish 8 to 10 steps. Rovan and Hayward also found that temporal behavior, the "envelope" of a short burst of "tactile sound," communicates more clearly to a musical performer than simple variation of frequency does.

Another way to deal with the problem of mode in a VE is by using voice commands to broaden the input interface (Thiébaux 1997; Nejad 1998). But command-based input, by voice or by keyboard, has inherently high latency. Similarly, increasing the input bandwidth of a synthetic instrument's controls necessarily adds extra degrees of freedom to the instrument.

3.3.3.3 Technologies for motion tracking

Several technologies are available for tracking the three-dimensional position and orientation of objects. Pattern recognition of video camera observations is still difficult. Moreover, it has inherently high latency because at most 30 frames per second can be measured. Magnetic and optical tracking have low latency (down to 4 msec) and high accuracy (a few mm, over a range of several meters) but cost tens of thousands of dollars.

Systems based on ultrasound require unobstructed paths between frame of reference and sensors (ultrasound transmitters and microphones, respectively). To ensure that enough paths are unobstructed and to overcome interference from ambient sound and reflected ultrasound, often several transmitters and receivers are used. For example, the 1999 touring show *L'Universe* by the Flying Karamazov Brothers uses six transmitters around the perimeter of the stage; the older GAMS system uses four transmitters and a "wand" of microphones attached to the performer. This extra equipment increases both cost and complexity. Like any ultrasound system, GAMS has moderate latency due to the finite speed of sound in air, 30 msec over a 6 meter square. Ultrasound also has fairly low spatial resolution compared to other technologies. Interestingly, GAMS can trade off spatial resolution with latency. At one extreme, 3 mm accuracy is possible with latency around 200 msec; at the other, 8 cm at the practical 30 msec limit (Bauer and Foss 1992). The inexpensive Flying Mouse and head tracker made in 1998 by Logitech claimed 20 msec latency and 0.1 mm accuracy, but experience proved it to be too slow, inaccurate, and susceptible to interference for serious use in musical applications or VE's.

In 1998 InterSense began to sell a hybrid ultrasound/gyroscopic motion tracker. It has very low latency and high accuracy, but the tracking elements are too heavy for use with smaller musical instruments, and the cost can be well over $10000. InterSense also sells a gyroscopic orientation-only tracker, the InterTrax2, for about $1000. This device could work well with musical instruments because of its low 4 msec latency and light weight, about 40 grams plus cable (InterSense 2001).

Micromachined accelerometers became inexpensive in 1997 when they were first mass-produced for automotive applications (air bag triggering and tilt sensing for alarms). From measured linear acceleration it is simple to compute linear velocity and hence linear distance, *i.e.*, position. Latency is typically on the order of 30 msec; a single-axis device costs about $30. For motion tracking, the most convenient device is triaxial, a single 30-gram unit measuring acceleration in three orthogonal directions. But since freely held orchestral instruments can rotate as well as translate, measuring their position requires measuring their orientation as well. Two spatially separated triaxial devices still fail to measure the sense of a rotation along the axis connecting the two, so a total of three spatially separated devices are required. The high cost (over $1000) and unwieldy geometrical calculations required to convert nine linear accelerations into reliable measurements of position and orientation suggest that we make some simplifying assumptions. Since the instrument is tethered to a stationary synthesizer, we assume a more or less stationary performer. If we take as fixed the performer's chin or pelvis, a single triaxial accelerometer mounted on the scroll of the violin can measure the orientation of the instrument as the scroll moves through a roughly spherical shell about this fixed point. Similarly a single accelerometer mounted on the frog of the bow suffices to determine its lateral position with respect to the bridge (as in the R-Bow (Trueman 1998)), as well as its orientation. Another complication is that accelerometers cannot distinguish between acceleration and tilt when in a gravitational field (like that of Earth). Correcting for this makes even approximate tracking of position and orientation difficult as slight errors begin to accumulate over the duration of a performance.

Fortunately one magnetic tracker is affordable. The SpacePad made by Ascension Technologies tracks the position of several sensors (30 grams, 15×15×25 mm) in a hemisphere of radius about 2.5 m. A two-sensor system costs about $1400. Latency of this system is 17 msec; resolution is about 1 cm in position and 3 degrees in orientation. Its weakness is the possibly large distortion of the magnetic field by nearby ferrous objects (steel in floor or walls) and cathode ray tubes. Such distortion is common to all magnetic trackers and is typically corrected by measuring the distortion at a number of points and estimating a function to invert this distortion.[129] This is called calibrating the motion tracker. Livingston (1997) proved that the distortion in measurement of orientation is a function of orientation as well as position (despite the manufacturer's claims); my own measurements confirm that the difference between actual and reported values of position and orientation varies for both as either is varied. (Livingston (2000) reports that no systematic relationship has yet been discovered between actual and reported orientations. With very restricted change of position, orientation could be calibrated in isolation; but in such a case, a gyroscopically based sensor of only orientation would be both more accurate and less expensive than the SpacePad.[130]) The distortion as measured by the SpacePad is therefore 6-dimensional, so calibration entails measuring the difference between actual and measured position and orientation on a 6-dimensional grid. Even with a coarse spacing of 30 cm in a 150 cm cube and 45 degrees of spacing in azimuth, elevation, and roll, this results in $5{\times}5{\times}5 \times 8{\times}5{\times}8 = 40000$ points to measure. Concert performance requires a setup of minutes, not days, so conventional calibration is impracticable.

Since a much faster calibration was needed, I investigated a nonstandard procedure. First, roll is omitted as this is quite awkward to control with the violin; this reduces the number of dimensions of the space to five. For each of the remaining dimensions (x, y, z, azimuth, elevation) the violin is waved through a path which holds that dimension at a constant value but goes through the full range of the other dimensions; this is done for several values of that dimension. For example, x is held constant to 100 cm while waving the violin through the plane of possible y and z values and simultaneously varying the violin's orientation and elevation; or the violin always points west while varying its elevation and moving through the whole x-y-z cube. Since data is measured at 60 points per second, this method produces actual/measured pairs of points thousands of times faster than conventional point-at-a-time measurement. A complete set of measurements is acceptably fast, taking about 15 minutes of violent sensor waving counting rest breaks.[131] The tradeoff with this calibration is much faster collection of data points but less data at each point: for each point only one actual dimension is known, not five. To

---

[129] Calibration cannot correct for time-varying distortions, of course. For several months, the resolution of my SpacePad was reduced to about 8 cm while the laboratory housing the eviolin suffered from an oscillating external magnetic field with a period of about two seconds. I suspected machinery in one of the adjacent rooms; one day the oscillation vanished suddenly, before I could verify this hypothesis.

[130] The Ascension 3D-Bird, several InterSense products, and some consumer equipment measure orientation and not position. Orientation can actually be measured outdoors with battery-powered flux-gate compasses and clinometers; the best of these cost about $300–700, and have sub-degree accuracy and 25 msec latency. I have combined this orientation measurement with GPS position measurement to create convincing headphone-based outdoor soundscapes (Goudeseune and Kaczmarski 2001).

[131] The sensor need not be attached to the violin for these measurements. For fixed sensor positions, I measured a typical difference of 2 mm and a worst-case difference of 5 mm with the violin present or absent. Ferrous metal in the strings and machined tuning pegs causes this difference. The sensor could even be mounted on the end of a lightweight 1-meter pole for faster calibration, but then orientation is difficult to control.

estimate the distortion of one of these five dimensions, say azimuth, a five-dimensional array is created (about 30×30×30×10×5 elements; this barely fits into memory on current desktop computers). Most values are uninitialized; points for which the azimuth value is known are initialized to that value. So if azimuth is held constant for values of north, south, east, and west, the array will have certain cells with values 0, $\pi/2$, π, and $3\pi/2$. Then each cell in the array is assigned the average value of its neighbors; this assignment is repeated until the values of the cells are not changing much, in the spirit of finding the equilibrium temperature of a plate constrained to certain temperatures at a few points. (For azimuth, unlike the other dimensions, special code is needed to handle the fact that 0 and 2π represent the same value. Roll, had it been measured, also has this special property.)

In practice, this abbreviated technique of measurement unfortunately does not produce enough information to adequately model the magnetic field distortion, particularly the distortions of azimuth and elevation.[132] The distortion is corrected exactly at the measured points, but the amount of data collected during calibration is insufficient to correct distortion adequately between these points. Some intermediate method which holds constant more than one but fewer than four dimensions while varying the others might be accurate enough. At present, the eviolin uses a simplified calibration which corrects distortion of only position, not orientation, holding one positional dimension (x, y, or z) constant while varying the other two. This calibration takes about five minutes.[133]

If the eviolin and bow are held conventionally, azimuth can be estimated from the position of the violin and bow: the normal to the vertical plane containing both of them, averaged over an interval of one to two seconds, indicates in which direction the neck of the violin points. (If the bow hand is too close or too far from the violin for accurate measurement, simply assume that azimuth has not changed since the last reliable measurement.) Elevation can also be estimated: unless the performer does deep knee bends while keeping the violin horizontal, elevation can be almost directly inferred from vertical height. Orientation could also be measured by a hybrid system, a separate device which measures orientation but not position. Ascension Technologies and others make such devices using either accelerometers or gyroscopes. They tend to be heavy, ferrous and expensive, thereby decreasing performer comfort, further distorting the magnetic field in a sensitive position near the violin's position sensor, and impoverishing the instrument builder. For these reasons the hybrid approach is less promising at the moment than simply estimating azimuth from the location of the two position sensors.

---

[132] For example, pointing the uncalibrated eviolin north and walking around typically produces azimuth reports anywhere from southwest through north and over to east. Spinning the uncalibrated eviolin in place, the reported values for north, south, east and west are not at right angles to each other.

No matter how position is measured, discrete gestures like dips and wobbles can be parsed from the continuous spatial data streams by deriving velocity and acceleration from the raw position data. (This gets into the fields of pattern recognition, filtering, and signal processing.) Of course, the null gesture must be recognized when the performer is just moving around with no intent of making a gesture. Also, the size, speed, and perhaps other attributes of the gesture should be computed; noting only the presence or absence of a gesture could be accomplished more with a button than a motion tracker.

Hybrid techniques are most promising for tracking both position and pitch, by analogy with how humans use multiple senses ("multimodal interfaces"). Intersense's hybrid of inertial gyroscopics and ultrasonic triangulation effectively tracks position, while the good performance of the PPPT trumpet pitch tracker is due to its measurement of valve position as adjunct to frequency estimation (Cook 1993).

Hybrid systems also show promise in gesture recognition. Sharma et al. (1999) have combined off-the-shelf speech recognition with hand position tracking from a video camera to build systems like a map one can point at and query, "Where is the closest parking lot to here?" or "How many people live in this area?" In this map system, recognizing hand gestures (as indicating a point, a line, or the perimeter of an area) improves the system's recognition of speech; also, speech recognition improves hand gesture recognition. In both cases, this improved accuracy of gesture recognition is attributed to the extra context available from the other modality; the extra context helps disambiguate unclear instances.

---

[133] Orientation without position, pointing the instrument in different directions without walking around, is slightly more compatible with reading from a music stand, avoids the hazard of tripping over cables, and allows somewhat faster gestures than walking. In such a scenario, using an orientation-only tracking device would work best.

4. The Eviolin

This chapter discusses the practical design of a synthetic instrument, the eviolin, based on the theory developed in chapters two and three. Numerous solo studies and compositions using it are detailed here, while a longer chamber work including it, *COALS*, is considered in the subsequent chapter.

Recall from the analysis of both orchestral and synthetic instruments in chapter two how input gestures, output sounds, and the mappings connecting them interact. Traditionally the main thrust of control has aimed at subtle yet flexible pitch manipulation. Amplitude control needs to be even more flexible but less subtle, so it is often driven by a derivative control. Timbral control is the most neglected, for historical and notational reasons to be sure, but also because variation of pitch is heard so much more readily—involuntarily, even—than most timbral variations.

Recall also that however the controls of an instrument are apportioned to drive its acoustic parameters, input gestures need to be matched to the performer. In times of slower technological development, inefficient or uncomfortable gestures could vanish of their own accord over the years, but now we need to consider ergonomics explicitly: posture; the inertia, accuracy, and speed of various effectors (arms, fingers, mouth); and the accuracy and temporal sensitivity of various sensors (proprioception, hearing pitch, hearing loudness, seeing hue, seeing spatial movement, and so on).

Also a number of heuristics suggest how to map gestures to sound, for instance following familiar physical patterns of energy transfer and excitation. By using interpolation schemes, mappings can be built up from individual connections between input control values and output acoustic values. Even after all this care has gone into designing a new synthetic instrument, it is quite appropriate for it to require some time to learn how to play well, perhaps not years but certainly months.

A synthetic instrument can be motivated from asking what sounds a given controller might produce, or equally well from how one might perform a given family of software-synthesized sounds. Ryan (1992) calls synthetic instrument design the putting of "physical handles on phantom models," discovering which controls ("handles") work well with mappings into a synthesizer (abstract "models" of sound).

4.1 Description and analysis of the eviolin

4.1.1 Augmenting an orchestral instrument

This synthetic instrument, the eviolin, is based on an orchestral instrument in order to take advantage of existing performance skill, to avoid waiting years for a performer to develop demonstrable skill on something entirely new. I wanted this instrument to work well in chamber music. Much work has already gone into keyboard controllers, and besides, chamber music has historically worked better with several monophonic instruments than several polyphonic instruments. In all periods, the repertoire for small ensemble has only a small fraction of works for more than one clavier, guitar, *etc*. So this suggests not keyboards but some other orchestral instrument.

Beginning with an easily muted orchestral instrument narrows down the choice of existing base instrument to some member of the string family. Bowed strings have a greater range of gestures and timbres than plucked strings, and thus give a wider range of potential experiments. But bowed strings may also be plucked, offering both sustained and decaying (percussive) timbres. Narrowing it down to one instrument of this family, the (electric) violin, is justified for three reasons: it costs less than its larger cousins; its large pool of performers facilitates finding accomplished players who also have technological aptitude; and its performer can walk around while playing, thereby allowing a wide repertoire of spatial gestures by means of motion tracking.

4.1.1.1 Extensibility to other synthetic instruments

Strings and keyboards are by no means the only orchestral instruments augmentable to new synthetic instruments. Such augmentation has been done with the trombone (Bromwich 1997), percussion (manufacturers such as Zendrum and Kat), woodwinds (Yamaha WX-7), and even several blown instruments at once (Cook 1992). Not even remotely orchestral is the design of yet other synthetic instruments or controllers, for example the Hands (Waisvisz 1985), aXi0 (Cariou 1994), dance-driven controllers, and the Meta-Instrument (de Laubier 1998). The techniques of both instrument design and composition illustrated by the eviolin certainly apply to other instruments.

Though numerous, the technological aspects of the eviolin are individually well understood problems and can be individually applied to other synthetic instruments. Pitch tracking devices are commercially successful. Motion tracking is possible through various technologies; if used for attributes of sound which tolerate higher latencies, even inexpensive consumer video equipment can be pressed into its service. Real-time software synthesis of sound is becoming common with the advent of faster computers and open-source software design. The craft of the instrument designer is one of integration, first making these disparate technologies all work at the same time and then making them feel like a single unit, a body rather than a collection of limbs. The better this second job is done, the simpler the mental model the performer has of the resulting instrument.

4.1.1.2 Augmenting an instrument will not replace the original

How to turn one functional prototype into thousands of instruments played by a large community with a significant body of literature is a problem beyond the scope of this dissertation. The entrepreneurial skills required to commercialize a prototype have rarely been found in the same person as the musical and engineering skills required to design new instruments. As an introduction to this social problem, (Théberge 1997, 41–51) analyzes the difficulties inventors have had in turning prototypes of musical instruments into products over the last few decades. Still, the eviolin's "one-box" design with off-the-shelf hardware is more easily duplicated than that of hardware-intensive synthetic instruments.

Almost no musical instruments developed in the twentieth century were accepted by the musical community at large. This can be attributed to the rarity of instrument designers who are also good entrepreneurs, as well as to the reluctance of composers and performers to invest time in an unproved instrument (leading, of course, to its perpetual unprovedness). Synthetic instruments are particularly vulnerable to this, as they depend on the rapidly evolving electronics industry (Théberge 1997, 41–71). It is prudent, therefore, to let a performer transfer years of developed skill on an orchestral instrument to a new synthetic one, instead of learning an entirely foreign interface. As much or as little as the synthetic and orchestral instruments have corresponding acoustic behavior, so much should their controls and mappings correspond.

4.1.2 Electromechanical construction

The electric violin is a five-string model made by Jensen Musical Instruments (figure 20). It is tuned as a standard violin, with the addition of a viola's low C string. The instrument has no resonant body so it is impervious to acoustic feedback even at high loudspeaker volume. The machined tuning pegs are mounted near the chinrest, not at the scroll; this places weight comfortably near the instrument's support instead of at the end of a cantilever.[134] The support for the Kun shoulder rest is specially made from aluminum instead of steel, to reduce the amount of ferrous metal distorting the magnetic field which measures the position of the violin.

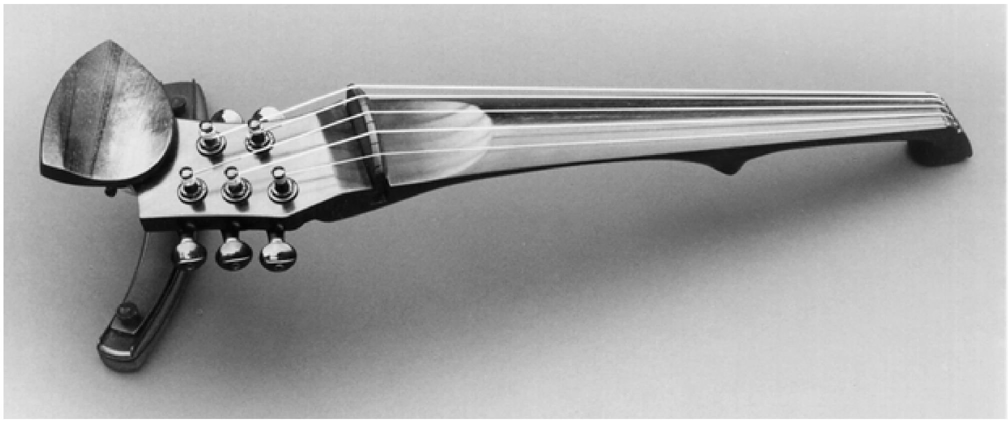


Figure 20. Electric five-string violin by Jensen Musical Instruments, used as the basis for the eviolin.

The position of the violin and bow are measured with a SpacePad tracker from Ascension Technologies. This consists of an antenna emitting a time-varying magnetic field and two sensors. The antenna is a collection of wires hexagonal in shape and about a meter across, held in a vertical plane by a collapsible wooden structure (figures 21 and 22). Its wooden construction and the positioning of the antenna midway between floor and ceiling is again an effort to reduce distortion of the magnetic field by metal in the building's structure. The position sensors are approximately 1×1×2 cm in size and weigh about 30 grams. One sensor is attached with double-sided tape to the bottom of the violin's neck (in the middle of the violin's back, if it had a body).[135] This is adequately distant from the ferrous metal in the tuning pegs at the chinrest, while still not impeding motion of the left hand along the neck.[136] For tracking the position of the bow, I first mounted the other position sensor on the back of the third finger of the right hand, like

[134] Practised violinists will want to note these details. The neck is solid like that of a guitar; the instrument overall weighs a little more than an acoustic violin. The Barbera pickup bridge feels flatter than a regular bridge, facilitating triple-stopping but also facilitating accidental bowing of an adjacent string when single-stopping. The lack of a body means that the palm of the left hand cannot nudge the instrument back onto the shoulder if it slips off. (Mark Wood's V-neck violin solves this last problem by supporting the instrument on breastbone and shoulder instead of using a conventional chin rest.)

[135] In figure 19, the middle of the neck shows a protrusion corresponding to the end of the body on an acoustic violin. This provides a tactile point of reference for the left thumb. Some violinists appreciated the motion sensor, mounted about halfway between this protrusion and the bridge, as a second point of reference when playing in very high positions.

[136] The builder of the violin, Eric Jensen, contributed advice on sensor placement and cable routing.

a ring. This avoided an extra mechanical coupling between bow and hand, and also avoided making the bow heavier. The third finger was best since it is never lifted from the bow and sits opposite the thumb. This gave good data but became uncomfortable with extended use.[137] So instead of this, the sensor is attached with velcro to the leather palm of a fingerless bicycle glove. Although the glove introduces some flexibility to the coupling and hence produces noisier data, it is quite comfortable: the glove was originally designed to be worn for hours on end, and it distributes force from the sensor's weight and cable throughout the hand instead of concentrating it at a point.

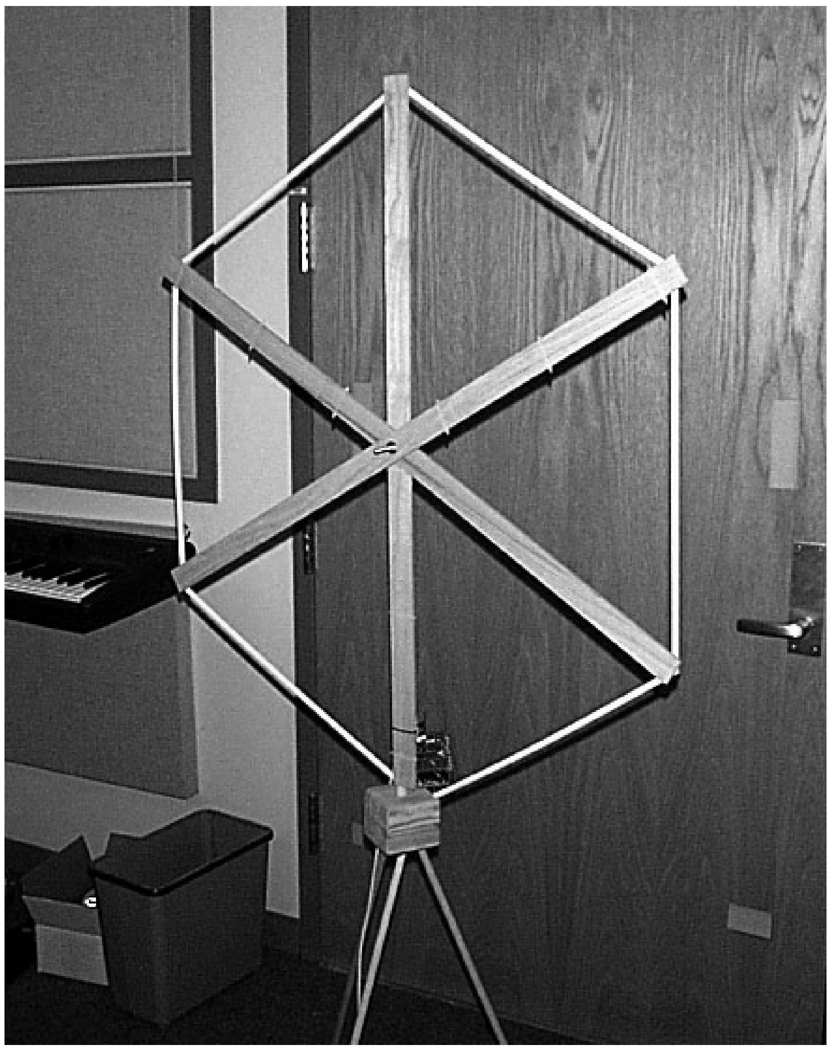


Figure 21. The motion-tracking antenna.

[137] Violinists Eric Rynes and particularly Chad Peiper evaluated my bow-sensor designs.

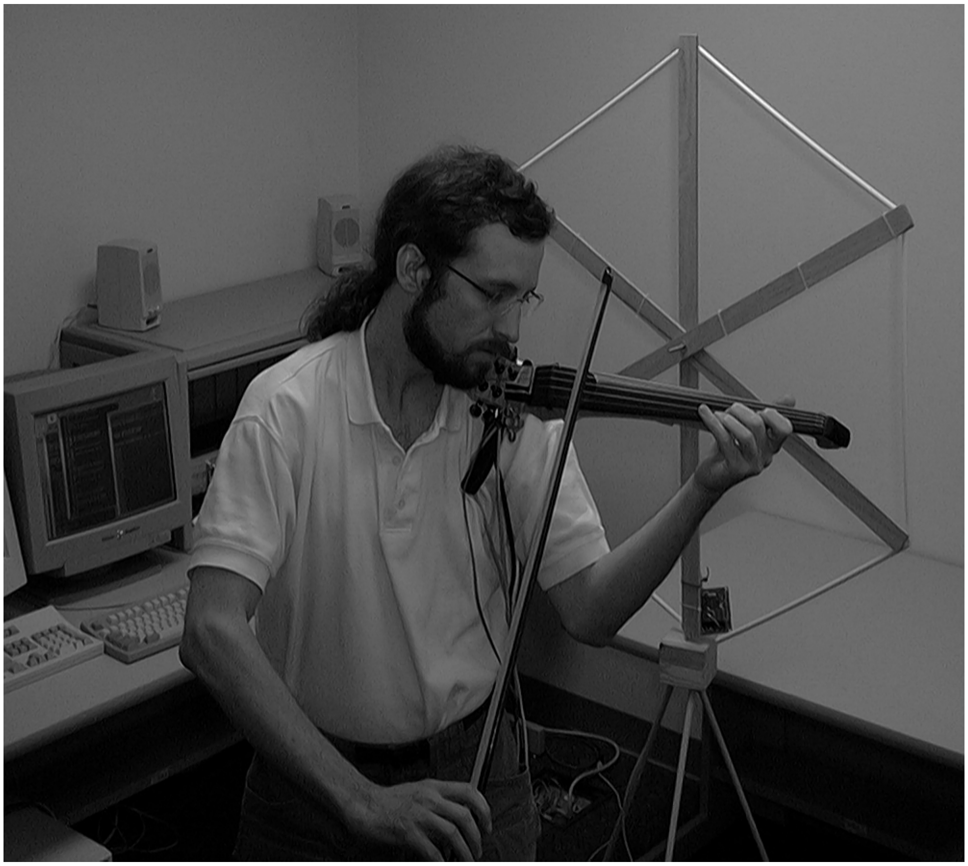


Figure 22. The eviolin and motion-tracking antenna in a laboratory setup. Photograph by William Sherman.

Simple visual feedback is provided by a collection of red, green, and blue LEDs mounted in a light diffuser, a small box about 1×2×4 cm weighing a few grams.[138] This display has only 3 msec latency, though it could theoretically be mere nanoseconds. The display is fixed to the strings above the bridge with elastic bands, so it returns to its position if it suffers a mild shock. The display sits about 8 cm from the violinist's left eye, in the lower left corner of vision. This is too close to focus on, which is actually an advantage since it displays not spatial patterns but only variations of brightness and full-spectrum hue. In principle this display can be driven by any parameter, for instance shifting to red if a note is flat, blue if sharp. But in practice it is most useful when it provides feedback for motion tracking, for instance to alert the eviolinist that he is approaching and crossing a boundary. The simple corresponding visual display here is a gradually increasing brightness followed by a sudden decrease. This feedback facilitates the

[138] Peter Sleggs helped design the discrete circuit which drives this display from the PC's parallel port.

synchronization of positional gestures with other gestures like bow changes or fingering changes.[139] Some versions of the *ondes martenot* use similar nonacoustic feedback to increase pitch accuracy. Transverse hand position drives pitch continuously, but a discrete (nonfunctional) keyboard helps the player aim for particular pitches.

Though simple, the LED display feels natural. It is qualitatively different for the performer to get feedback from the artifact in his hands instead of from a large display some distance away. Performers still prefer to think of their instruments as objects, not as distributed networks of heterogeneous equipment. If some visual feedback requires a larger display, it can of course be used together with the LED; the large display serves well for high-latency high-bandwidth feedback, the LED for low-latency low-bandwidth feedback with a small number of degrees of freedom (particularly for positional feedback).

The tether connecting the performer to the computer is noticeably heavier and more fragile than a standard electric guitar cord. The weight of the tether could be halved by building a single shielded cable to replace the audio, two motion-sensor, and LED cables currently held together with cable ties every 50 cm. For now, a belt-hook a few feet down from the violin end of the tether puts the brunt of the weight on the hip instead of on the violin and bow hand.

The acoustic vibrations of the strings are transduced by the bridge into an electrical signal. This signal is strengthened by a preamplifier and then sent to the input of a computer running VSS software. This software analyzes the signal, reporting things such as its pitch, loudness, and spectral centroid. The pitch tracking is derived from a public-domain software tracker called fiddle~ (Puckette, Apel, and Zicarelli 1998); for violin waveforms, fiddle~ has latency comparable to that of commercial hardware pitch trackers (Yoo and Fujinaga 1999).[140] The software also interfaces to the motion tracking hardware, implements all the mappings from controls to sound which make this a synthetic instrument, and computes the final sound which is sent to amplifier and loudspeakers (figure 23). The signal sent to the loudspeakers can also be stored directly on the computer's hard disk as a sound file, for purposes of rehearsal or noise-free recording.[141]

---

[139] Humans respond more quickly to haptic than visual stimuli (Sheridan and Ferrell 1981, 262), so piezoelectric buzzers in the palm of a hand or inside a watchband on the wrist might work well too.

[140] Software tracking of pitch directly in VSS has another advantage: VSS can simultaneously track other acoustic parameters like percussive attacks, brightness, and vibrato with no additional hardware and no task switching of the CPU.

[141] During performance the signal is recorded into a large preallocated block of memory (100 MB stores roughly ten minutes of 16-bit 2-channel sound at 44.1 kHz). Only afterwards is this block copied to disk, since writing to disk during performance occasionally causes audio dropouts at the low latency the eviolin requires.

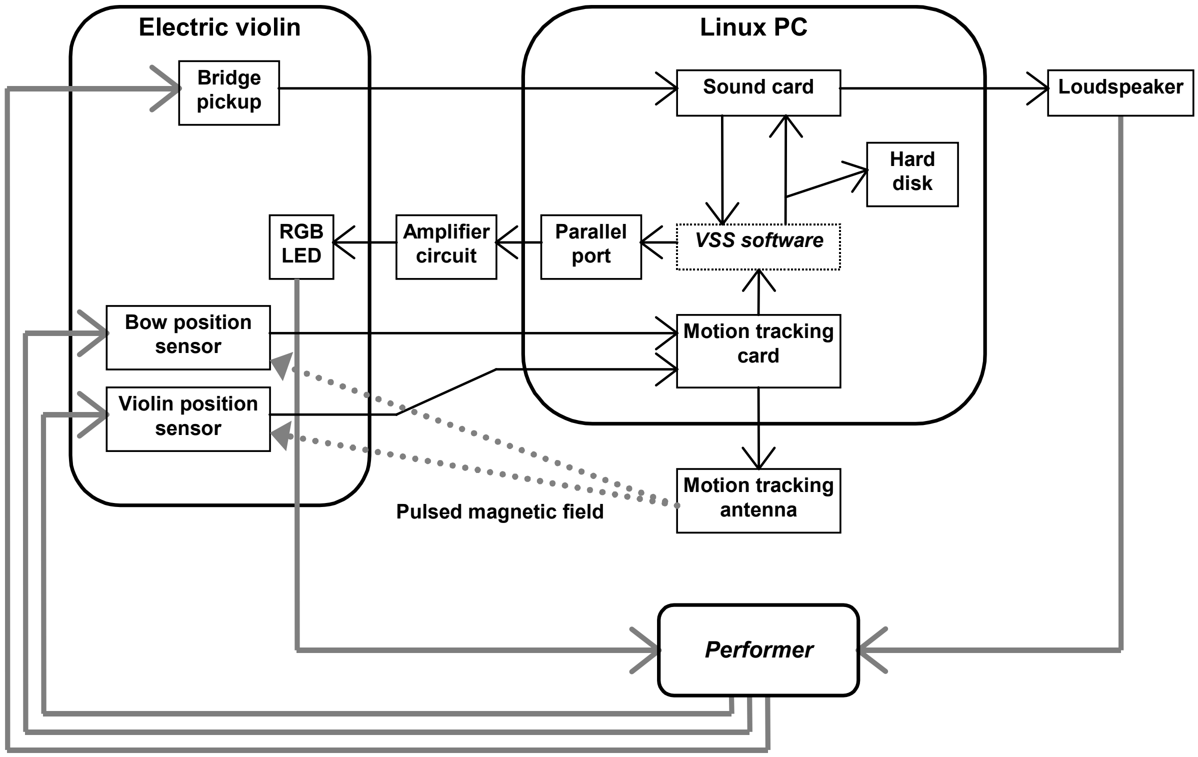


Figure 23. Block diagram of the eviolin.

4.1.2.1 Attenuating direct sound

One reason for choosing the electric violin was its relative lack of direct sound compared to other orchestral instruments. But in playing it "unplugged" I found that the direct sound was actually quite loud.[142] During demonstrations in a fairly small laboratory (3 by 5 meters), guests have asked for headphones so the synthesized sound could mask the sound emanating directly from the strings. The direct sound may have little to do with the synthesized sound: for example, I programmed the eviolin so that its strings sound as if they are tuned a major ninth apart, and fingered semitones sound as whole tones, fingered scales as arpeggios. In this case the synthesized sound is generally at a different pitch from the direct sound and masking does not happen naturally. This is quite an issue for the performer, whose ear is right next to the strings.

Muting the (very thick) bridge might attenuate the direct sound, but would not improve matters because the synthesized sound would be attenuated even more (assuming amplitude maps to amplitude). On an acoustic violin the bridge transmits string vibration to the body, but on an electric violin the bridge arguably *is* the body, and damping its vibrations affects the pickup more than the strings. I considered applying cloth to the fingerboard, using alternate string materials, muting the strings at the nut, and muting the strings above the bridge; Trueman (2000) agrees with me that these are not all that effective. Strings different enough to sound appreciably quieter might have several disadvantages, too: a different coefficient of friction with the fingertips feels peculiar to a violinist performing a shift of position; the strings may go out of tune more quickly; and most seriously, with sustained use thin strings may injure the player's fingers (Rynes 1999).

Rynes (1999) also argues *for* keeping the strings audible. Even with a classical violin, the violinist uses the direct sound to check pitch, hear shifts of position, set the bow on the strings, check a difficult high note with an almost inaudible left-hand pizzicato, and so on. If the eviolinist needs to hear more of the synthesized sound, additional loudness can come from a headphone on the right ear, the ear pointing away from the strings. Also, if the eviolin is 5 meters back from the edge of the stage, the direct sound is practically inaudible to the audience in the third row (although other performers on stage still hear the direct sound). This is apparent in concert, but not in the laboratory.

Most simply, the performer can reduce direct sound by bowing less vigorously and turning up the volume on the eviolin's preamplifier. Even as simple a modification as amplification changes an instrument enough that the performer does well to change certain gestures—in this case, "sawing" with heavy bow pressure. More dynamic range is available on electric violins by varying bow speed than by varying bow downforce (Jensen 1998).

4.1.2.2 Design of the motion-tracking antenna

The specification given by Ascension Technologies for constructing the antenna requires the arrangement of wires following three loops which outline a hexagon with edges 48 cm long. In the interest of portability, I also required that it be collapsible, lightweight, and fairly rugged (able to withstand a few bumps and falls). My first design was constructed like a hexagonal stunt kite. The wires were taped to a sheet of plastic with a pocket at each of the six corners, into which snapped three thin wooden dowels along the main diagonals of the hexagon. Two struts coming from the center propped up the kite at an angle, sitting on the floor and facing the eviolinist. This met all my requirements—it weighed under 100 grams and rolled up into a bundle 95 cm long and 4 cm in diameter—but it performed poorly. I discovered that the motion tracking was more accurate if this antenna was propped up a few feet above the floor (still facing the violinist); this was because the floor (and ceiling) of the university buildings I

[142] For truly inaudible direct sound, an electric lute might be necessary!

was playing in contained considerable ferrous metal in the form of steel rebar and air ducts, enough to distort the magnetic field emitted by the antenna. The distance between this metal and the antenna needed to be increased; it does not matter that the metal in the floor was on the opposite side of the antenna as the motion sensors were. The distortion is reduced if the receiver is much closer to the antenna than the metal is, but this could not happen while the antenna was mere inches from steel in the floor. No thorough mathematical model of this distortion has been published, but it has been shown repeatedly that even in near-ideal conditions, measurement error increases with the fourth power of distance; future magnetic trackers with much greater range are therefore not expected.

So the redesigned antenna had to contain very little ferrous metal; also, its hexagon had to be far from both floor and ceiling, but close (on average) to the violin.[143] Despite this, I wanted it to have few components, be quick to set up at a concert, be easy to build, and cost a reasonable amount when compared to the other components of the eviolin.

This antenna is made almost entirely of wood (figure 21). Overall cost is about $25, weight is under 1 kg, and travelling size is 110×15×10 cm. The three propellor-like slats are oak; the tripod base and the little support from the cubical center to the center of the hexagon are made of pine dowels. The three oak slats are mounted on a stainless-steel bolt with a 1 cm ring at its head, a specialty yachting item. After loosening the bolt's wing nut, the oak slats swivel on the bolt and collapse for transport. The ring at the bolt head snugly surrounds the tapered tip of the topmost pine dowel. For neatness, the wires run through plastic tubing from a plumbing supply store (plastic hose costs less but does not straighten out completely).

The cleverest part of the construction is the wooden block between the hexagon and the tripod, made from two short lengths of two-by-four glued together. Vertically into its top is a hole into which the hexagon support is inserted. Around this hole and slightly intersecting it, three more cylindrical holes are drilled at angles into the bottom of the block. The tripod dowels go into these holes; but as the holes overlap, the dowels overlap slightly and have notches cut out of them. The whole assembly therefore fits together snugly like a Chinese wooden puzzle and cannot fall apart accidentally; the dowels can be inserted and removed from the block only in a certain order. This is done with no metal fasteners whatever.

---

[143] To avoid "standing waves" from a time-varying magnetic field parallel with large masses of ferrous metal, the hardware engineer of the original CAVE suggested that the antenna not be parallel with walls or floor (Dawe 2000). But I failed to measure a change of accuracy when propping up the antenna at different angles at a single position, so I made the hexagon of the redesigned antenna vertical for better stability.

4.1.2.3 Visual feedback module

The LED module mounted above the bridge renders 3+3+2 bits of red/green/blue color intensity, driven from the 8-bit parallel port of the host computer. For each color, its multiple bits create summed currents through a small resistor network. Brightness could instead have been controlled by flickering the LEDs invisibly fast (over 1 kHz) and varying the duty cycle, but I rejected this as it would have meant either yet another real-time task for the computer or a fairly complicated circuit to convert an 8-bit value into a duration. The overall design is deliberately simple because elaborations like greater color depth or arrays of multiple modules are generally not cost-effective when compared to consumer video units such as small LCD screens or eyepieces. Such elaborations also tend to make the tether between violin and computer quite heavy, compared to a single video cable. Although lightweight, fiber optics is too fragile for use in the tether. A short bundle of fiber optics could have been used on the violin itself, though: the LEDs could then be mounted underneath the violin, and a bundle of about a hundred strands, stiffened by a single-core aluminum wire, could curl from underneath to a conveniently visible position. This bundle would also blend the colors more uniformly than a conventional light diffuser.

A high-resolution display (granted, any display with more than one pixel is high resolution!) offers greater flexibility of visual representation; it can even display multiple representations at once. Such a display, if taken full advantage of, demands a significant amount of attention from the performer. By contrast, a single-pixel display communicates at an almost subconscious level, like a pianist feels the sides of the black keys to determine hand position when playing chords. Of course high-resolution displays can work like this also, if restricted to low-level visual attributes like brightness, hue, saturation, speed of blinking (between two hues, between bright and dim), and size and tilt of simple iconic images. A wall of information, particularly text or numbers, may not be that useful even while practising, never mind performing: such detailed information can only be adequately attended to when not actually playing. The performer might play a one-minute passage with only low-level visual feedback, and then sit back and review the detailed information collected during that interval, like the conductor of an ensemble (or the teacher of a student) gives detailed verbal instructions only after stopping the music.

4.1.2.4 Tetrahedral Loudspeaker

Trueman and Cook (1999) have eloquently demonstrated the advantages of spherical loudspeakers to conventional loudspeakers for instruments whose sound needs to blend with non-loudspeaker instruments: the loudspeaker can be located very near to the instrument's controls, and it radiates sound in all directions (and potentially differently in each direction). Such a system is appropriate for the eviolin, because it should be able to blend with orchestral instruments.

As no experience with such loudspeakers with fewer than twelve driving elements had been published, I decided to experiment with a four-driver system. The raw material is a surround-sound system intended for video games, the Sirocco Crossfire. This system consists of a five-channel amplifier, a subwoofer, and four small satellite speakers.[144] The satellites are epoxied to a block of wood shaped like two triangular prisms joined at right angles to each other, schematized in figure 24. (Epoxy and wood are used since ruggedness and lack of ferrous metal is a concern as with the antenna.) In this configuration the satellites point outwards like rays from the center of a tetrahedron to its four vertices; it may be easier to visualize this as rays to four nonadjacent vertices of a cube (see figure 25). This loudspeaker "molecule" sits on top of a 30 cm wooden column; the angle of the column's peaked top matches the 'V' gap between two satellites. The mating surface between column and satellite-molecule is covered with felt to prevent buzzing and vibration. The column rests on the subwoofer; a ten-conductor cable connects the whole assembly to the amplifier. The amplifier has both two- and four-channel inputs.

By conventional measurements of sound quality—frequency response, harmonic distortion, efficiency, *etc*.—this loudspeaker is more than adequate. Distortion becomes perceptible only at painfully loud levels. One concern with the present satellite-molecule is diffraction around its many edges; this can be reduced by filling in its notches to make it quasi-spherical or at least convex.

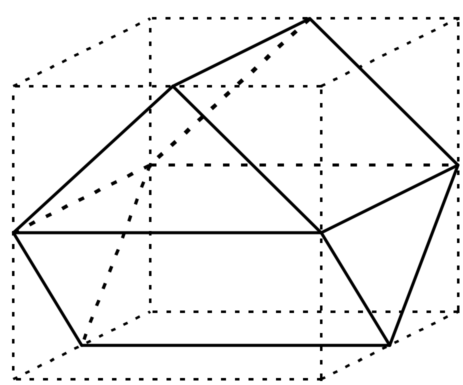


Figure 24. The wooden block connecting the four satellite speakers, and the 75 mm cube from which it was cut. The speakers are epoxied to the four square faces. The two upper square faces are at right angles to each other, as are the lower two.

[144] Since the eviolin can generate a full frequency spectrum, it would not suffice to have a loudspeaker array with only tweeters or only midrange drivers as is the case with Trueman and Cook's 12-driver spheres.

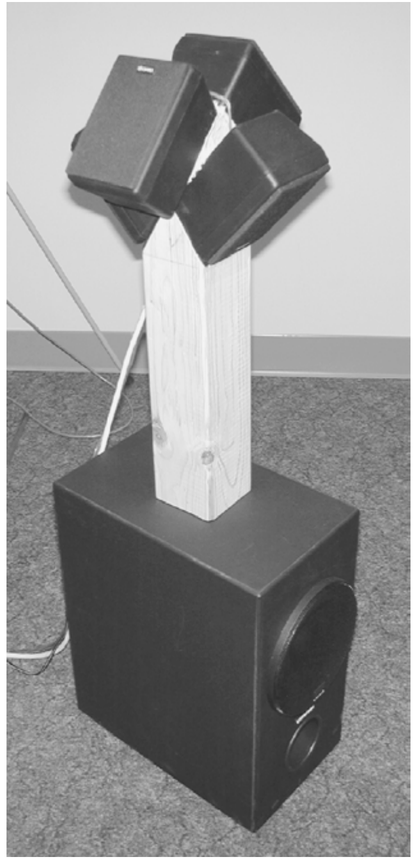


Figure 25. Loudspeaker configuration of the eviolin.

4.1.3 General controls

Configurations of the eviolin commonly include some controls derived from the bridge pickup and some from motion tracking. The former includes pitch, amplitude, and spectral brightness data. The latter is more involved. The raw data is the position (latitude, longitude, height) and orientation (azimuth, elevation, roll) of the sensors mounted on violin and bow, relative to the antenna; but performers think more naturally in terms of derived data like playing at the tip or at the frog, *i.e.*, the distance from frog of the bow to the bridge of the violin. Other derived data are the position and speed of the bow relative to the violin and the violin relative to the floor. Measuring the bow relative to the floor is less useful because all of a violinist's training aims at how the bow interacts with the violin. (When the bow is not exciting the string, though, the floor may be a better reference frame than the presumably nonstationary violin.) Separate x-y-z components of position are useful controls for the performer, but this is not so with velocity: it is difficult to think at the same time of high z-velocity, low x-velocity, and high y-velocity. Performers more easily think of only an overall speed, or a speed of a certain magnitude in a certain direction.[145]

For all the eviolin configurations in this chapter, tracked pitch and amplitude drive output pitch and amplitude respectively unless otherwise stated. This obviously natural mapping of input to output is easy for a violinist to learn quickly, so only special cases depart from it.

4.1.3.1 Pitch tracking

The extraction of frequency-domain data from a set of bowed strings is a less well defined problem than it at first appears. More accurate pitch trackers generally have greater latency. Hybrid pitch trackers consume more resources but may do better than their individual components. Various strategies exist for tracking the pitch of each string individually. Pitch tracking can even be avoided entirely by keeping the signal in the time domain.

Since strings continuing to resonate (particularly open ones) can occasionally confuse the monophonic pitch tracker of the eviolin, Chad Peiper, Guy Garnett and I experimented with various mutes on the strings. The old standby of weaving a folded dollar bill through the strings above the bridge reduces this ringing somewhat. More extreme modifications of the strings near the nut like stuffing various densities of foam rubber under them or even wrapping rubber bands *over* them damp only open strings, not stopped strings. Reducing the inertia of the strings themselves by changing string material is counterproductive, for the same reasons given against this as a way to attenuate direct sound. These "prepared violin" tricks reduce ringing on the scale of a half second, but on the scale of a tenth of a second (which is long enough to confuse the pitch tracker) the reduction is not so dramatic. The most practical solution is simply preferring fingerings which avoid open strings.

Both to work around this ringing-string problem and to allow double-stopping, Guy Garnett, Tim Johnson and I investigated how well pitch tracking worked if we independently tracked the signal from each string. I internally rewired the bridge of the eviolin to put out separate signals from the individual strings, and built a cable which terminates in five unbalanced quarter-inch plugs. (The cable also has a switch to revert to monophonic output.) These five plugs connect to the microphone-level inputs of a mixing console, which sends the five separate signals over an ADAT fiber-optic cable to a Macintosh running Opcode Max.

---

[145] Visual artists who deal with technology experience similar mismatches. Just like x/y/z velocity components are convenient for computers and awkward for eviolinists, amounts of red, green, and blue are convenient for software but have little to do with how humans cognize color.

As an initial experiment Tim Johnson and I connected the signal from the A and E strings to two fiddle~ objects, and sent the MIDI output of the fiddle~ objects to a Yamaha Disklavier (a MIDI-playable grand piano).[146] Trying different values for the fiddle~ control parameters, double-stopping on the A and E strings works reasonably well. A single fiddle~ object quickly tracks pitches above middle C, but at the expense of including glitchy triggers of higher harmonics. No bow attack is entirely free of these glitches; it is quite fun to play thicker or thinner rolled chords on the Disklavier by more or less aggressively attacking the string with the bow. Playing *tratto* (with too much pressure) causes a non-stop barrage of such glitches. Rejecting spurious fundamentals which really are higher partials is possible only by winnowing the output of fiddle~, not by adjusting its control parameters. Unfortunately this increases latency. A compromise, lower latency with fewer glitches, may be possible by second-guessing the output of fiddle~: if it has suggested fundamentals of 800, 1600, and 1200 Hz in the last 10 msec, then immediately report 400 Hz as the true fundamental even before the raw output of fiddle~ has settled down.

The polyphonically wired eviolin exhibits crosstalk. Measuring the direct output of all strings while bowing only one, the unbowed strings output the same signal as the bowed string. The replicated signals are slightly attenuated, typically 9 to 20 dB lower. Their signal strength is largely independent of how the unbowed strings are damped, even when damped extremely and creatively as described above, so it must be due to internal mechanical coupling in the bridge.[147]

Besides the Disklavier, we also drove a pair of simple additive synthesis instrument from this pair of fiddle~ objects tracking the A and E strings. In this configuration, playing a single pitch over and over again results in successive notes with enormously varying amplitudes. This may be due to constructive and destructive interference in the simple spectrum of additive synthesis: the phase of the two synthesized sounds will change from attack to attack due to slight variations in when the two pitch trackers lock on to a new note. More complex synthesizers suffer less from this phase difference, but chorusing effects are still present.

To better isolate these signals and avoid such duplication, each of the five pitch trackers could have a limited range of acceptable fundamentals, for example from the lowest frequency of the string to a twelfth

[146] The Max program would often crash when running more than one fiddle~ object, and sometimes add distortion when the pitch was not a multiple of 256 samples. By trying out different computers, we isolated these problem to the ASIO driver for the 8-channel sound card on our Macintosh.

[147] Eric Jensen chose the Barbera bridge for his electric violins in part for its *high* inter-string isolation. Jensen actually refused to sell me a more widely known MIDI violin bridge, reasoning that it would only earn him an unhappy customer. He is currently developing his own polyphonic bridge.

above that. Even so, multiple pitch trackers will react to a single note sometimes: in this example, the open E string will also be picked up by the A- and D-string trackers. A voting scheme might handle this case: if several trackers report values within a quarter tone of each other, winnow out all but the loudest. Still, some special cases would need to be designed to avoid acoustic artifacts during note attacks. Deliberately playing the same pitch on two strings (with vibrato), as is sometimes notated for increased volume, would be the most difficult test case for such a voting scheme, or playing open A together with a glissando on the D string past that A.

Another musician working with electric violins suggests eliminating crosstalk by using optical pick-ups or by having a mechanically separate bridge for each string (Jehan 2000). He also outlines a software-only approach to a solution, as follows. For each excited string, and for each of the other resonant strings, measure the amount of crosstalk over the full frequency range. This produces 5×4 = 20 functions of amplitude versus frequency. Modify fiddle~ to take five inputs instead of one; once spectral peaks have been found by the FFT, evaluate every possible combination of bowed strings with respect to these functions to see which hypothesis of bowed/unbowed strings best explains the five spectra measured. (Brute force suffices for this: since bowed strings must be adjacent, only 5+4+3+2 = 14 combinations need evaluating, from single- to quadruple-stopping.) But for multiple-stopping this does not avoid the thorny problem of polyphonic pitch tracking, as a set of mechanically separate bridges does; also, a software model of physically possible left hand shapes is required to guess which pitch comes from which string.

4.1.3.1.1 Latency

The worst-case latency of the eviolin, the longest time from the onset of a bow stroke to the onset of the corresponding sound, is about 65 msec. Technically, this is latency of amplitude; other latencies can be defined by measuring the duration between other gestures and their acoustic effects, for example the larger latency between a change of bow speed (inferred from successive bow positions) and some acoustic dimension driven by this speed.

Of these 65 msec, operating system latency contributes 20 msec; I have measured it as low as 16 msec.[148] This is measured by sending a sound into the line-in jack of the soundcard on the Linux PC, copying this sound in software to the line-out jack, and recording both these signals, the input and output, with the stereo line input of a second computer. By visually examining the left and right channels of this stereo waveform in a sound editor, it is straightforward to measure the delay separating the two signals. Patches now available for the most recent kernel for Linux, version 2.4, promise latency under 3 msec (Morton 2001).

The remaining 45 msec of measured latency comes from pitch tracking, a figure consistent with that reported by Yoo and Fujinaga (1999). Clean bow attacks above middle C are faster, typically 35 msec. (The same measurement technique applies: record both the violin signal and the Linux PC's output signal using the stereo line input of a second computer.) Several strategies can further reduce this amount.

If very fast passagework is demanded, the direct sound (perhaps processed) can be mixed in with the synthesized sounds driven by pitch tracking. The performer then has at least some acoustic feedback as fast as with an acoustic violin. The pitch-tracked sounds can come later, 45 msec later or (with a soft attack) even half a second later. The difficulty with such an approach is how to perceptually fuse two distinct sounds, direct and synthesized.

Hybrid pitch tracking is also possible. Though the fiddle~ pitch tracker is quite robust and accurate, simpler algorithms can be faster. For example, an autocorrelation-based tracker would work well if the bottom register of the violin was not tracked. If fundamentals are recognized in the range 400 to 3200 Hz with 1/8-tone accuracy, on a signal decimated from 44.1 kHz to 22.05 kHz the worst-case latency is about 11 msec (4.5 periods of a 400 Hz waveform), and the measured load on a 450 MHz Pentium CPU is about 1%, low enough to run continually.[149] When fiddle~ detects a transient (a sudden increase in energy), it could immediately query the autocorrelator for its rough opinion of what the pitch is, and keep using that approximation until fiddle~ itself had settled on an accurate pitch value. If the two trackers disagreed noticeably, crossfading their two values over 10 msec might improve the result.

FFT-based trackers like fiddle~ have poor accuracy when constrained to have low latency, because their sample window is then so short. Of course there are other ways to measure pitch, like zero-crossing detection and filter bank techniques such as Fundamental Period Measurement (Kuhn 1990). These are generally less tolerant of attack transient noise and weak fundamentals than FFT's and autocorrelators. Since they may also require much tweaking of coefficients before they work satisfactorily, I have not considered these further.

---

[148] Trying to reduce latency by using separate sound cards for input and output produced mysterious buzzing sounds. This was probably because one sound card acted as master clock and the other slowly drifted out of synchronization (each card had its own crystal for a time reference).

[149] The input to the autocorrelator should be high-pass filtered, to remove potentially confusing energy below 400 Hz.

The physical parameters which give rise to pitch can be sensed directly, particularly finger position via resistive strips, or optical or ultrasonic measurement (Paradiso 1999c, 1998, 1997). Such devices are difficult to generalize to more than one orchestral instrument, though. They also can be fooled by "bending" pitch (bowing forcefully or overblowing) or by playing harmonics by touching a node for only a moment. Using only the signal itself as input to a pitch detector is robust in that it assumes nothing about how the instrument is played (or even what the instrument is).

4.1.3.2 Processing the direct sound with resonance models

Pitch tracking latency can be eliminated by eliminating pitch tracking, in particular eliminating the conversion of a time-domain signal into the frequency domain and back. No unpredictable lag occurs if we keep the signal from the violin's bridge in the time domain and modify it only with traditional processors like filters and reverberators. On the other hand, such a restriction begins to threaten the claim of the eviolin to be a synthetic instrument, one where software can flexibly define the connections between input gesture and output sound. Granted, the other data streams of amplitude and position can still modify the audio processing by means of general-purpose software, but the instrument loses the flexibility of remapping pitch or mapping pitch to non-pitch parameters.

The eviolin is one of six instruments in my chamber composition *COALS* (2001), analyzed in detail in the next chapter. In the context of small chamber ensembles, very low latency of amplitude (under 10 msec) is required to maintain a tight sense of rhythm (particularly tutti attacks), so here Guy Garnett and I decided to use an advanced sound processor controlled by the eviolin's tracked elements. The sound processor is the resonance model from the software package CNMAT Additive Synthesis Tools (CAST) which runs on a Macintosh. Motion tracking data and pitch tracking data are collected by VSS on the Linux PC as usual, but instead of being sent to an audio synthesizer in VSS itself they are sent over Ethernet to the Macintosh. The audio signal from the electric violin is split, sent both to the PC for relatively high latency pitch tracking and to the line input of the Macintosh for passing through the resonators (figure 26).

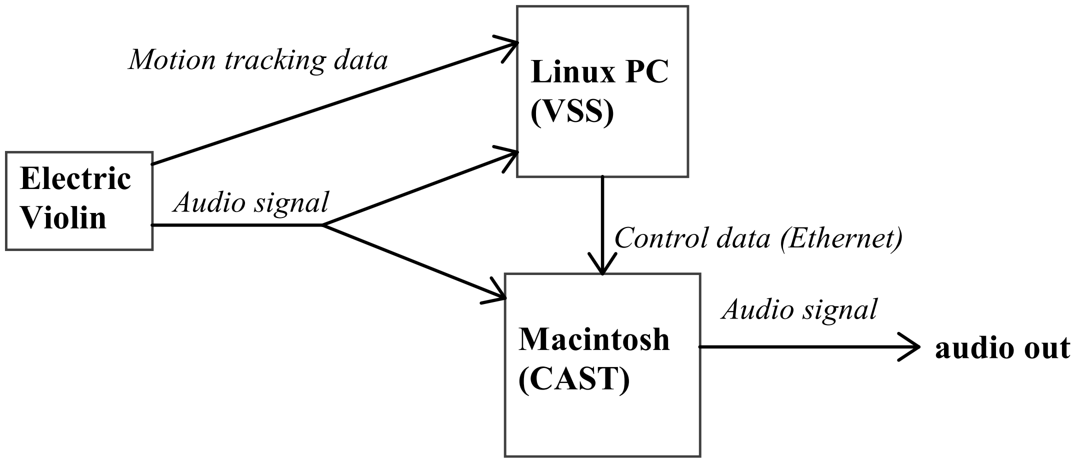


Figure 26. Configuration of the eviolin for resonance processing with CAST.

The CAST software grew out of research by (Barrière, Potard, and Baisnée 1985; Potard, Baisnée, and Barrière 1986). Although the resonance model is implemented as an efficient bank of filters, its output sounds quite similar to additive synthesis of exponentially decaying sinusoids, or partials. Each partial has its own frequency, initial amplitude and decay rate; decay rate can also be interpreted as inversely proportional to the bandwidth of an IIR filter (Freed 2000a, 2000b). Tim Johnson and I built up a resonance space for CAST by hand, placing particular timbres at floor positions by means of the simplicial interpolator implemented in VSS.

In figure 26, VSS acts as a server of violin data to the resonator client. This client/server protocol is implemented in OpenSound Control (OSC), a layer built on top of the UDP Ethernet protocol (Wright 2000). The server waits for a handshaking message `/ViolinControl/Connect` from the client; since UDP is connectionless, this initial message contains the address of the client to which the server will send data. (Only one client at a time can connect to the server.) The client then tells the server which data it wants and how often, by sending a `Param` message for each data type it wants (figure 27). For each of these `Param` messages, the argument indicates how many milliseconds should elapse between messages. If the argument is zero, that datum will be sent at its maximum rate (as fast as VSS itself gets updates from the SpacePad or fiddle~). If the argument is negative, that data stream is turned off. The client also sends `Event` messages for any data it wants which are discrete events, as opposed to the continuous data streams of the `Param` messages. Finally the client sends `/ViolinControl/Enable` to start receiving data from the server.

Message	***Argument***	***Function***
`Connect`	string	host name of client computer
`Disconnect`	(none)	
`Enable`	boolean	enable/disable reporting of data to client
`Param/Pitch`	float	rate at which to report MIDI pitch number
`Param/Frequency`	float	rate at which to report frequency
`Param/Amplitude`	float	rate at which to report linear amplitude
`Param/X`	float	rate at which to report x-coordinate of violin
`Param/Y`	float	rate at which to report y-coordinate of violin
`Param/Z`	float	rate at which to report z-coordinate of violin
`Param/Mix/0`	float	rate of reporting amplitude of 0^{th} resonator
`Param/Mix/1`	float	rate of reporting amplitude of 1^{st} resonator
`Param/Mix/...`	...	...
`Param/Mix/9`	float	rate of reporting amplitude of 9^{th} resonator
`Event/NoteOn`	boolean	enable/disable reporting of note onsets
`Event/NoteOff`	boolean	enable/disable reporting of note terminations

Figure 27. Messages understood by the eviolin data server.

From the server the client then receives messages listed in figure 28, each at the particular update rate which it requested. During reception of data, the client can at any time change rates (perhaps as part of an adaptive-loading scheme) or enable or disable reporting of individual parameters. Reporting stops when the client sends another `/ViolinControl/Enable` message, with an argument of false instead of true. When the client sends the `/ViolinControl/Disconnect` message, the connection is broken and the server is ready to communicate with another client.

Message	*Argument*	*Function*
`Param/Pitch`	float	MIDI pitch num. (60.02 = 2 cents above middle C)
`Param/Frequency`	float	frequency, in Hz
`Param/Amplitude`	float	linear amplitude, in the range 0 to 1
`Param/X`	float	x-coordinate of violin, in the range -1 to 1
`Param/Y`	float	y-coordinate of violin, in the range -1 to 1
`Param/Z`	float	z-coordinate of violin, in the range -1 to 1
`Param/Mix/0`	float	linear amplitude of 0^{th} resonator (0 to 1)
`Param/Mix/1`	float	linear amplitude of 1^{st} resonator
`Param/Mix/...`	...	...
`Param/Mix/9`	float	linear amplitude of 9^{th} resonator
`Event/NoteOn`	(none)	a note began
`Event/NoteOff`	(none)	a note ended

Figure 28. Messages sent by the eviolin data server.

Gain control proved to be a difficult problem. A resonator might be accidentally chosen so that the violin signal has little or no energy in the frequency bands of the resonances. The eviolinist then feels the instrument refuse to get loud no matter how hard the bow saws across the strings, quite a frustrating experience. Conversely, if a partial in the violin signal suddenly lines up with a high-Q resonance (either because the violinist changed pitch, or because the resonance changed due to movement of the violin), then the output signal suddenly becomes very loud. To avoid the second situation, such narrow resonances are simply disallowed. The first situation, where most frequencies present in the violin signal are blocked, is more involved. Since the violin signal is filtered into many narrow quasi-sinusoids, we could insert enormous amplification factors at various points in the Max patches (figure 29) without amplifying background noise and without clipping. For adjusting values for these amplifiers, I recorded a few dozen sound files of raw violin signal with a variety of pitches, loudnesses, and bowing techniques. This adjusting was more easily done when we could simply trigger these sound files by means of the buffer~ and play~ objects defined in Max.

As another experiment in making narrow resonators usable, I used the height of the eviolin to cross-fade between raw violin signal and broadband noise. The amplitude of the noise was scaled by the amplitude of the violin sound. So the lower the violin dipped, the less pitched was the sound going into the resonator; but the likelihood of at least some frequencies making it through the resonator would approach certainty.

The `Param/Mix` data indicate the relative strengths of up to ten resonators, to be interpolated between (creating a composite resonance, frequency band by frequency band). The simplicial interpolator implemented in VSS ensures that the ten amplitudes are all in the range 0 to 1 and that they sum to unity. Then each resonator can be assigned to a point in space, and the up to ten-way crossfade automatically happens as the violin moves around these points.

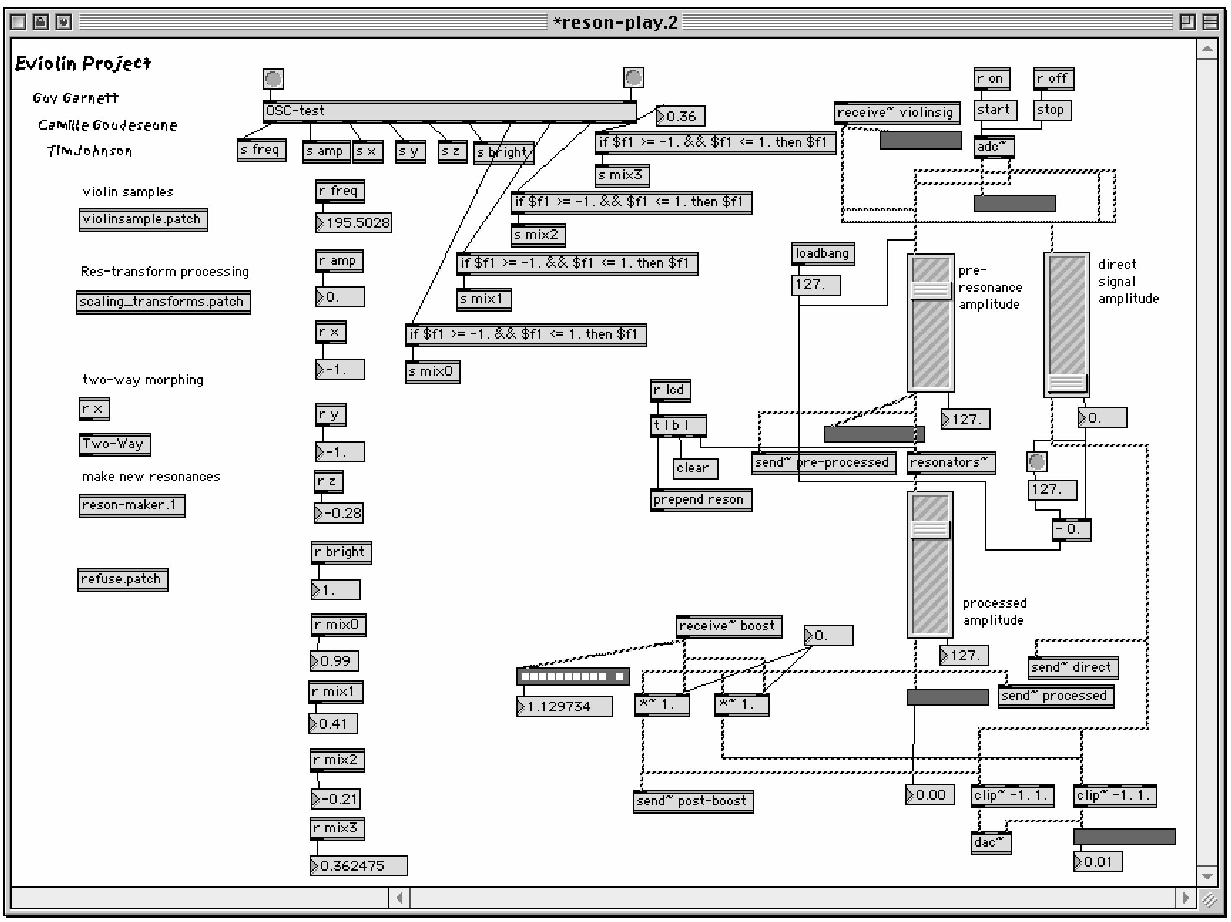
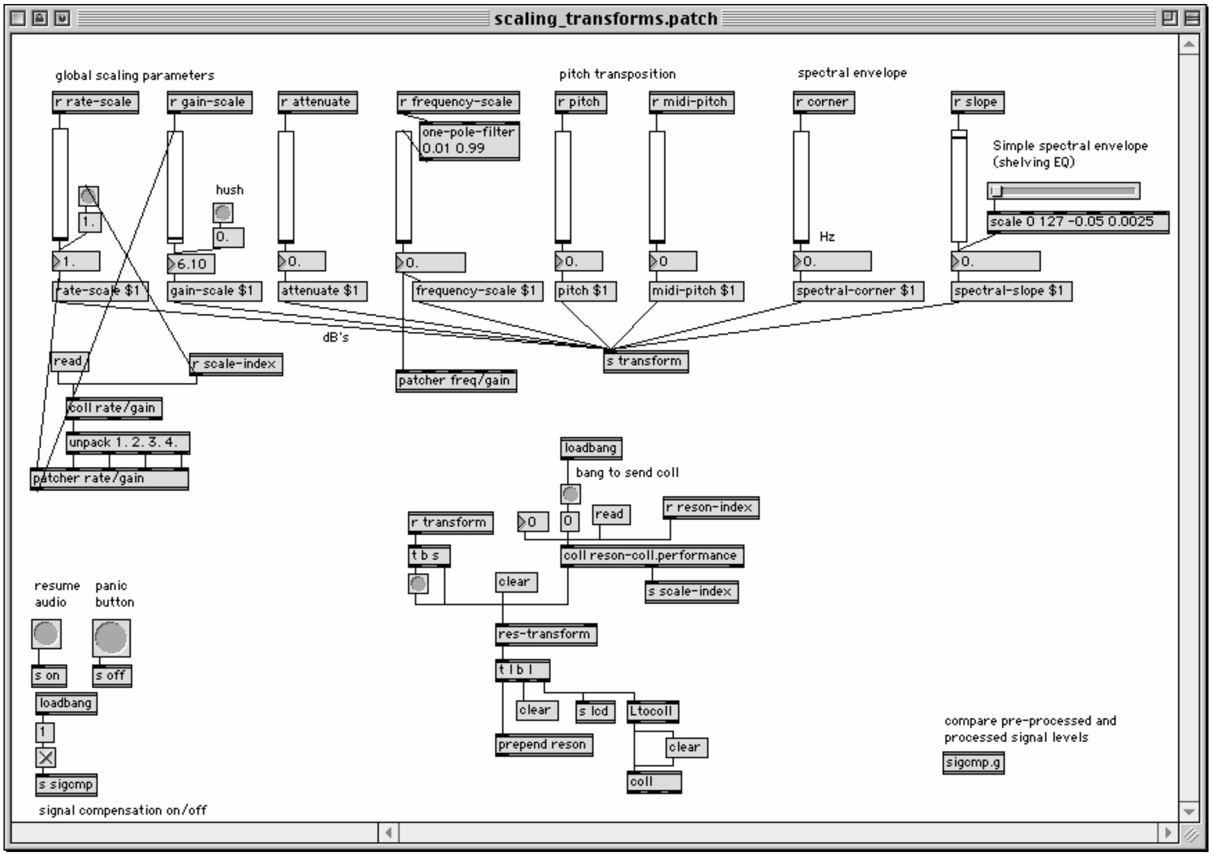


Figure 29. The Opcode Max "patches" which use data collected from the OSC server to control the CAST resonators.



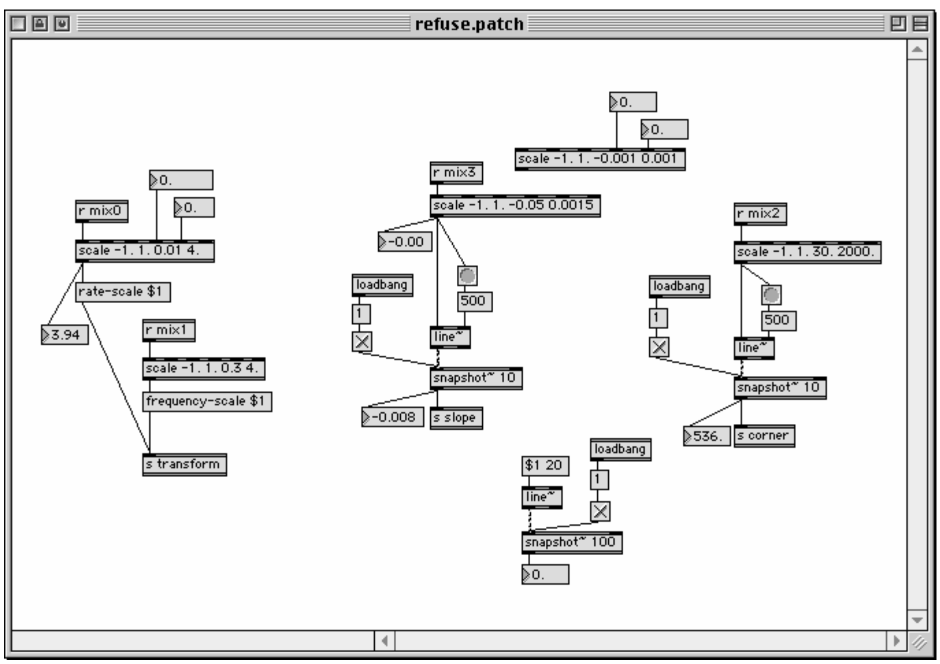


Figure 29, continued. The Opcode Max “patches” which use data collected from the OSC server to control the CAST resonators.

For the composition *COALS* the eviolin needs two independent controls of timbre, one driving spectral brightness and the other driving spectral richness (a few strong resonances, or many weak ones). This composition uses only one resonance at a time, since interpolating between resonators is tricky. In general, such interpolation sounds more muddled than working with individual resonances, either one at a time or simply crossfaded in amplitude. After having crafted a few resonators, smearing them together is like carefully preparing steak and potatoes but serving them as differently proportioned dabs of puree. But interpolation can work in a special case, when all the resonators are chosen to be compatible—members of a family of resonators generated by a small number of index parameters, all with the same number of resonances, all with fairly similar characteristics.

4.1.3.3 Tracking of individual partials

Many musicians prefer to use pitch tracking for triggering events or following scores, instead of driving a synthesizer from every tracked note.

It is really a vast oversimplification to represent an audio signal with only two parameters, pitch and amplitude. At the other extreme, representing an audio signal by its raw sample data gives far more parameters than can be used to control a synthesizer. Even the raw frequency content is too much data, too far from the performer's mental model of the sound. It is possible to track things less rigorous than pitch *per se*, which can be both accurate enough and fast enough to feel good to the performer. One such intermediate representation is a collection of a few dozen spectral peaks or partials.

Partial tracking is implemented by using only the first processing stage of fiddle~. This stage produces a list of frequency peaks which a later stage normally converts into an estimate of a single fundamental; partial tracking dispenses with this estimation and simply reports lists of frequency peaks one frame at a time. Successive frames are converted into connected tracks of nearby peaks with the algorithm used by the open-source Loris analysis/resynthesis package (Fitz 2001; Fitz, Haken, and Christensen 2000). Unlike conventional pitch tracking, partial tracking adds only a few msec to raw operating system latency; it also handles double- and triple-stopping because no single fundamental is computed. Correctness of the partial tracker was verified by sending these peaks directly to an additive synthesizer which reconstructed (most of) the electric violin's sound. Energetic attacks like pizzicato, however, cause a cascade of high partials rippling down to the fundamental in about a tenth of a second. This cascading is greatly reduced by moving energy from higher to lower partials. This energy redistribution is done heuristically, without estimating a fundamental frequency (which would reduce to conventional pitch tracking, after all). As an example, if partials are present at 2401 Hz and 400 Hz, since 2401 is close to an integer multiple of 400, the heuristic reduces the amplitude of the former to zero and increases the amplitude of the latter. The overall reassignment of amplitudes conserves total energy. This heuristic does not go so far as to introduce a "missing fundamental" or missing lower partials, but merely redistributes energy already present in the spectrum. Its only side effect is occasional artifacts when double-stopping.

4.1.3.4 Spectral brightness

The fiddle~ pitch tracker of the eviolin is extended to also report dimensionless spectral centroid. This value is defined as

$$C = \sum_n f_n a_n \Big/ f_1 \sum_n a_n$$

where $n \geq 1$ ranges over all partials (the case $n = 1$ corresponds to the fundamental frequency), f_n is the frequency of the n^{th} partial, and a_n is its amplitude. For a pure sine tone, $a_n = 0$ for all $n > 1$ and thus $C = 1$. Higher values of C indicate relatively higher energy in the upper partials.

Controlling brightness with only sul ponticello / sul tasto playing is difficult. For notes played extremely sul pont., fiddle~ just bumps the fundamental up an octave and reduces the reported spectral brightness accordingly.

The spectral centroid strongly discriminates between an open string which is bowed and one which is left to vibrate and decay. When the bow leaves the string, the brightness suddenly drops a large amount and then smoothly descends to 1. (Stopped strings decay much more quickly.) The spectral centroid also strongly discriminates harmonics, both open and fingered, which are much closer to sine tones than conventional notes. It also distinguishes the same pitch played on different strings (high up on the G string, or back in first position on the A string). A bright steel E string always produces high values of C.

If the identity map from spectral brightness to spectral brightness is desired for a given synthesis algorithm, this can be implemented in real time by measuring the spectral centroid of both input and output and adjusting the output's brightness control to compensate. A simpler open-loop system, less accurate but also less prone to lag, measures both input and output brightness in the laboratory to produce a fixed compensation function for simple lookup during performance.

Reporting of spectral centroid could be augmented with reporting of spectral width ("richness"). A more involved analysis could produce two higher-latency controls, rate and depth of vibrato.

4.1.3.5 Tracking the motion of an instrumentalist

Three-dimensional space is isotropic to a geometer, but anisotropic to an instrumentalist in a gravitational field. The vertical direction is special in practice. Moving in a horizontal plane requires less muscular endurance and less attention than moving in tilted or vertical planes. This is attested to by the experiences of Bauer and Foss (1992) with the GAMS motion tracker, Fels and Hinton (1998) with GloveTalk II, and Paradiso (1999b) with the Magic Carpet. The GAMS tracker explicitly incorporates horizontal planes into its "space editor." Different regions of the space can have different attributes, but these regions are defined not in arbitrary three-dimensional terms but only as parts of horizontal slices of the space. While brass or woodwind players can easily alter the angle of their instruments, extended violin playing at elevations higher or lower than nominal becomes uncomfortable. Therefore I prefer to divide spatial gestures into two aspects, horizontal position (latitude and longitude) and momentary vertical gestures (speed, not position).

A study Guy Garnett wrote for the eviolin demonstrates this separation. Horizontal position drives two aspects of timbre, while vertical gestures drive the amount of reverberation. Reverberation is initially off. A slight dip of the eviolin adds a small amount, which remains after the dip. Lowering the violin farther adds correspondingly more reverberation. The amount corresponds to the greatest depth reached so far. Reverberation suddenly vanishes when the eviolin is raised well above nominal (in the study, typically at the end of a phrase).[150] The compromise here is more comfort for the eviolinist but no gradual reduction of reverberation. Of course gradual reduction could be added to this design by means of an extra control, perhaps a certain shake of the bow. This control could toggle dipping between gradual increasing and gradual decreasing; but then the eviolinist would have to remember which of two states the instrument is in.

This same study maps bow position, playing at the frog or at the tip, to pronounced and smooth attack transients respectively. This correspondence is natural to violinists, offering greater timbral variety with almost no extra demands on performer attention.

[150] Technically, the input gain to the reverberator is suddenly reduced to zero. The output gain is unchanged, so that when the eviolin is raised, any reverberant sound occurring at that moment still continues to reverberate.

4.1.4 Examples of eviolin configurations

4.1.4.1 Swaying Echoplex

The Swaying Echoplex configuration (figure 30) uses only orientation rather than position. This lets driven parameters change more quickly, since pointing the violin in a different direction is faster than taking a few steps across the floor. The index of modulation of the frequency-modulation synthesizer is driven by amplitude, to simulate the brighter timbre most instruments make when played loudly.

Elevation drives pitch discretely. When the eviolin points upwards at least 15 degrees above nominal, the output pitch is an octave higher; when below nominal, it is an octave lower. Hysteresis applies here, so octave jumps do not inadvertently occur during a single note: this octave switch is polled only when a note begins. The discomfort of playing in an upward-tilted position fits with the understood difficulty of playing in an extreme register (while still being easier than actually playing an octave higher). Both upward and downward tilt correspond intuitively with our idea of high and low notes, and need no explanation to the audience.

Azimuth drives the amplitude going into four echo units, delay lines with feedback. The echo durations are increasingly larger prime multiples of 50 msec. The overall decay time of each successive echo unit is about three times longer than the previous one. As the eviolin rotates to the right (flexor muscles contract), the longer echoes predominate; the hue of the LED module progresses from blue through green to red as the eviolin rotates from total absence of echoes through the shorter ones to the longest one. This formalizes how violinists sometimes sway in *espressivo* passages. A particularly dramatic long note (*e.g.*, one in extreme register) can then continue to ring like a pedal tone while shorter notes are played simultaneously. Sustained notes actually sound like reverberation, while quick notes sound like distinct echoes.

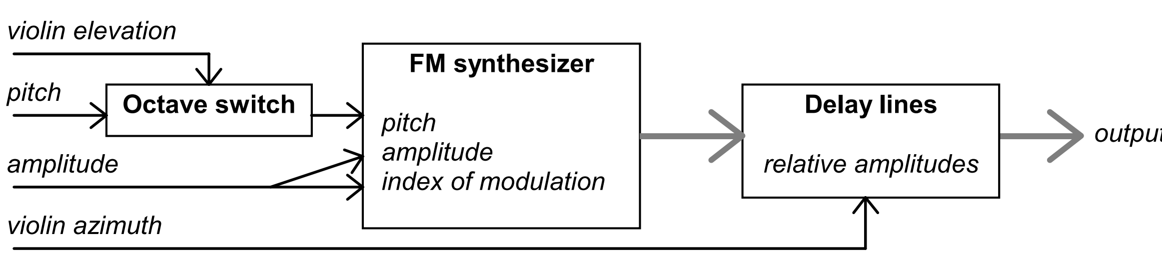


Figure 30. Driving graph of the Swaying Echoplex.

4.1.4.2 Adjustable sensitivity to intonation and dynamics

The synthesizer in this configuration is a physical model of a clarinet from the STK toolkit (Cook and Scavone 2000). Stage position affects pitch and amplitude. As the performer walks left or right, pitch is rounded to the nearest tempered semitone (as with a conventional valved clarinet) or left alone (allowing glissandos as with half-valving). Dynamics are reduced in range when downstage, always between *mp* and *mf*. Upstage they are exaggerated, so an input range of *pp* to *ff* becomes *ppppp* to *fff*.

The modifications of pitch and dynamics are driven here by position, not orientation, even though pointing the eviolin is faster than walking around. This is because nonstandard orientations (particularly elevation) become uncomfortable when maintained for a while, whereas the performer can stand anywhere indefinitely to sustain a particular instrumental behavior. These modifications typically occur slowly, not changing back and forth every few seconds.

4.1.4.3 Spatialized sound source

In this configuration, horizontal position of the eviolin drives the position of its sound source. As the performer walks around, the apparent position of the sound source (the unmodified signal from the bridge pickup of the eviolin, to keep things simple) moves correspondingly between two or four spatially separated speakers.

4.1.4.4 Pseudo-Theremin

To illustrate the flexibility of a software-based instrument, this configuration drives pitch from the distance between eviolin and bow-hand. The bow need not touch the strings, or even be held for that matter. Height of the bow hand drives amplitude. The timbre is a filtered square-wave; height of the eviolin controls the corner frequency of the low-pass filter.

As a final control, horizontal distance from the eviolin to the antenna drives amount of pitch correction. So depending on where the performer stands, a range from perfectly smooth glissandi through approximate scales to absolutely accurate scales is possible. Figure 31 summarizes this description.

Because of the lack of haptic feedback and the increased number of controls, for the novice this instrument is harder to play than a true Theremin, not easier. It demonstrates not only the flexibility but also the difficulty of learning a software-based instrument.

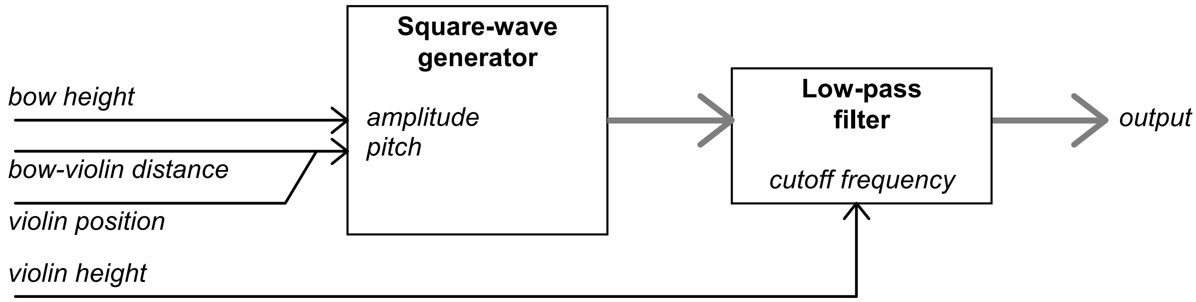


Figure 31. Driving graph of the Pseudo-Theremin.

4.1.4.5 Timbre-rover instruments

Of many timbre spaces built with the timbre rover and simplicial interpolator two are worth mentioning here, a drawbar organ and a vowel space. The first space is simple additive synthesis—the timbre rover computes the amplitudes of individual partials of a harmonic series. The second space controls nine parameters of a CHANT synthesizer (Rodet 1984): three formants, each with relative amplitude, center frequency, and filter bandwidth. Since subtle rather than dramatic changes of timbre work well here, and since one timbre might be sustained for an extended duration, these timbre spaces are controlled by horizontal position instead of orientation.

4.1.4.6 *Rarae Aves*

This configuration sprang from a desire to play in real time the sounds used in my 1999 tape piece *Rarae Aves* (“strange birds”). These double-modulator frequency modulation sounds have hardly any identifiable pitch but their spectrum is rich and highly varied. They explore the region around 10 to 25 Hz where individual beats blur perceptually with low pitches.

Double-modulator FM has a profusion of parameters to control: frequency ratios between carrier and modulator, between modulator and second modulator; two indices of modulation; and amount of feedback for each modulator. This sound is then sent through a low-pass and high-pass filter, so the gain for each of these need also be controlled. The timbre rover considers only frequency content and ignores time-variant behavior, so it would have applied poorly to this family of sounds where a single set of parameter values might produce a sound which changes enormously over a period of as long as one second. So instead I constructed the timbre space by hand, choosing individual points and passing them to a simplicial interpolator.

Only two dimensions impoverish this enormous range of sounds, so a conventional map of horizontal position to timbre is inadequate: either many interesting sounds are excluded or the space is incomprehensibly corrugated. Happily, tracked pitch is available as a third dimension: it is inappropriate for driving pitch itself for two reasons. First of all, the timbres are generally inharmonic so conventional pitch structures do not apply. Secondly, driving carrier frequency from tracked pitch to simply scale the frequency of the output signal would almost always destroy the interesting ambiguity between beats and very low frequencies. So the timbre of the instrument is driven by three controls, latitude, longitude, and pitch, each of which can be indefinitely held at a single value.

The instrument is actually implemented as a duet of synthesizers, not a single one as was done in the tape piece. For the same vector of input values, both synthesizers produce approximately the same timbre; they differ in that one has highly time-variant behavior while the other emits "purer" tones. Transverse position crossfades the output of these two synthesizers, so the performer can at least simply understand a connection between position and wobbliness of sound. This transverse line at each point punctures a timbre square. This square is perpendicular to the line, of course, and can be thought of as extending in the other horizontal direction and in the vertical direction; but the vertical direction here is not distance from the floor but tracked pitch (low C to high E, 129 Hz to 1320 Hz, scaled logarithmically). This is again because elevation can be varied only momentarily lest discomfort ensue. The position of the eviolin in this conceptual square is then sent to the simplicial interpolators driving each of the two synthesizers.

This instrument is more difficult for a violinist to learn than one which maps pitch to pitch, precisely because it is farther away from a conventional violin. At least tracked amplitude still drives output amplitude. One surprise is how odd the temporally varying timbres can feel. Since there is no control over the phase of these periodic variations, the eviolinist feels somewhat at the mercy of the algorithm. For instance, playing a very short sound of one desired timbre along this periodic variation is not possible: whatever the phase happens to be, that will be the sound which comes out.

4.1.4.7 Relative multislider with variable gain

In a timbre square laid out on the floor, the horizontal position of the eviolin acts like a computer mouse on a mouse pad. A movement northwards of distance A moves the "cursor" in the timbre square a distance B. Examples of this two-dimensional slider presented so far have only been absolute. But such a multislider can also be relative, varying its gain B/A (possibly differently along the two axes, even). This gain is driven directly from the violin's vertical position: nominal height has low gain for subtle control, while dipping and crouching increases gain so the performer can quickly reach a different timbral region and then stand erect to explore it with subtle nuance once more.

Since the multislider is relative, the violin position could attempt to move the cursor outside the timbre square. If this happens, the cursor keeps its last legitimate value within the square. As well, the onboard LED display blinks bright red to warn the performer of the overshoot so they can retreat a little towards the center of the square.

Playing the eviolin with adjustable gain in a rich timbre space such as that of *Rarae Aves* is fun and wildly expressive but easy to get lost in. Trying to play short musical scores, performers develop a strong desire to associate particular timbres with particular points in space. This may be so it is possible to aim for a particular remembered point in the timbre space as an open-loop gesture, not a closed-loop aiming and re-aiming. No amount of visual feedback helps this situation, since feedback inevitably closes the loop. The space is simply too general for performers to develop the typical open-loop gestures of a mastered instrument. Some simplification of the structure, perhaps dividing the timbre square into subsquares with uncrossable breaks between them, is necessary for mastery. A two-level multislider, with only a fixed high gain and a fixed low gain, is much easier to keep track of during performance. It is also easier to control: instead of a continuous control like height, a discrete control like a parsed bow-wiggle or crossing a vertical threshold suffices to toggle between the two gain values. At the low-gain level, a further constraint can be applied to the center point of the scaling: instead of placing the zoomed-in subsquare arbitrarily, it might round off to the nearest of a three-by-three array of fixed subsquares. The compromise with easier understanding is then giving up the ability to cross the boundaries between subsquares while at low gain.

Visual feedback helps the performer remember the current association between position and timbre. Most simply, momentary confirmation that gain has changed to a new value or warning that position has moved out of range are point-by-point reminders. Continuous visual feedback of this spatial relationship requires a visual display with spatial extension, not the monopixel display currently used by the eviolin.

4.1.4.8 The “diving underwater” metaphor

A useful mapping for the eviolin is from its position in the horizontal plane to timbre, since the performer can stand for an extended period at any position without becoming fatigued. Vertical position can add discontinuous gestures in timbre to such a mapping. Intuitively, as the performer walks around, timbre changes only continuously. If he crouches, going under the water’s surface as it were, then he can walk around without changing timbre; timbre instantly catches up upon resurfacing. (A loon paddles around for a while and then suddenly dives for minutes at a time, leaving only ripples marking its last sighting. Suddenly it reappears in a different place.)

To effectively control this system the performer must know exactly where the "water level" is, both when above and when below this level. This height is indicated by the rate of blinking of the LED. Hysteresis in the algorithm ensures that the LED blinks only when approaching the boundary, not when moving away from it (since when moving away, the performer needs no warning about it). The blinking is not overly distracting, because the performer does not stay underwater for a long time. The algorithm is symmetric. Approaching the boundary from above or from below, blink rate increases with proximity to the boundary; if the boundary is crossed, blinking suddenly stops; blinking remains stopped as long as the performer is still moving away from the boundary (figure 32).

```
fAbove = true;          // Are we above the boundary?
fJustCrossed = false;   // Did we just cross, and not yet turn back?

// This next fragment of code is executed repeatedly.  The values of
// fAbove and fJustCrossed are preserved between executions.
//
// y is current y-value;  yPrev is the previous value (a few msec ago).
// yBoundary is the y-value of the boundary.
// dyHyst is a small hysteresis value.

if yPrev < yBoundary and y >= yBoundary
    // we just crossed the boundary going up
    fAbove = true
    fJustCrossed = true
if yPrev > yBoundary and y <= yBoundary
    // we just crossed the boundary going down
    fAbove = false
    fJustCrossed = true
dy = y - yPrev
if (fAbove and (dy < -dyHyst)) or (not fAbove and (dy > dyHyst))
    // we just turned back towards the boundary
    fJustCrossed = false
if fJustCrossed
    Turn_off_blinking
else
    yDistance = abs(y - yBoundary)
    if yDistance > 3 inches
        Turn_off_blinking
    else
        As yDistance ranges from 3 to 0 inches,
            blink from 1 to 6 times per second.
```

Figure 32. Pseudocode for visual feedback while approaching a spatial boundary.

4.1.5 Conservation of energy

The most fundamental attribute of any sound is its amplitude: if its amplitude is zero, it has no other attributes to speak of. And since the amplitude of a vibration indicates its energy, our everyday experience of energy conservation laws applies here. With an orchestral instrument (or a car door for that matter), more energetic gestures produce louder sounds. With the eviolin, therefore, most applications drive the amplitude of the output signal primarily from the measured amplitude of the signal at the bridge.

One sometimes useful violation of this physical rule is to alter the rate of dissipation of energy, almost always slowing it down. The following non-stateless mapping is often useful: map input amplitude to output amplitude, subject to the condition that output amplitude can decrease no faster than some limit. This is conveniently implemented in the discrete world of digital audio by the pseudocode in figure 33, executed every few milliseconds.

```
// The value of amplOutPrev is preserved across executions of this code.
// Adjust the constant 0.997 to produce the desired decay rate.

amplOut = max ( amplIn, amplOutPrev * 0.997 )
amplOutPrev = amplOut
```

Figure 33. Pseudocode for clamped decay rate of signal amplitude.

I originally implemented this clamped decay for an early eviolin study. The clamping applied only to pitches below the open D string, so they could act as pedal tones (only a few seconds long, like the bass notes of a Chopin left-hand accompaniment) while the eviolinist continued to play other material.

4.1.6 Remapping pitch

If a synthesizer has a pitch input, that input is most intuitively driven directly from the tracked pitch of the eviolin. Tracked pitch can of course be mapped to non-pitch parameters, just as other things can drive a synthesizer's pitch input. A third possibility is to map pitch to pitch, but not with the identity map.

A simple practical example of this third possibility is correction of intonation. I implemented this pitch-to-pitch mapping as actually a family of mappings indexed by one parameter. At one extreme of the parameter's range, pitch is unchanged; at the other, the output pitch is the tempered semitone nearest the input pitch. Lateral position of the eviolin interpolates between these two extremes. This design was motivated by my own poor intonation: fast passages use more pitch correction, slow passages use less so that vibrato is possible. The LED is dark when the input pitch is exactly in tune, becomes redder as pitch falls flat from a tempered semitone, and becomes greener as pitch goes sharp. When pitch is exactly a quarter tone off, the LED is bright yellow (red plus green).

An even simpler example is an octave switch: the output pitch is one or more octaves above or below the input pitch, like octave couplers on a pipe organ. In one study I wrote for the eviolin, tilting the violin upwards causes pitches to sound an octave above normal. (To be precise, tilting the violin sounds an octave below normal; ordinarily, pitches sound *two* octaves below normal.) This switch is disabled for input pitches below the open E string, to avoid accidental triggering.

A more fanciful mapping is a linear expansion of input pitch. The open D string maps to itself; as the input rises a quarter tone, the output rises a semitone. Similarly, the output falls twice as fast for pitches below the fixed D. The end result is that six octaves are within the left hand's grasp, without going higher than first position on the fingerboard. Arpeggios can be played like scales. Previously unimaginable double stops such as fifteenths are easy (when polyphonic pitch tracking is present). The tradeoff is that the performer must be twice as accurate in intonation. Of course there is nothing mathematically special about doubling pitch range in this way. Expansions or contractions of factors other than two (or indeed of changing factors) are possible, but more complicated mappings are naturally more difficult for the performer to grasp intuitively.

Another family of pitch mappings takes one input pitch to a set of output pitches, letting a monophonic instrument play harmonies directly instead of by arpeggiation. The first such mapping I implemented went around the circle of fifths as pitch moved from one tempered semitone to the next. For example, an exact C mapped to that same C; as the input became slightly sharp, G and D were added; when the input got a quarter-tone sharp, all twelve pitch classes were present in various octaves. As the input continued to sharpen, it was reinterpreted as being a slightly flat C-sharp and pitch classes vanished from the sharp side, beginning with G sharp, ending with E, B, and F sharp, until only C-sharp was sounding.

This mapping required superhuman intonation skill, so I tried variants on four ideas from this implementation: (i) a simple pitch mapping for regularly spaced input values (the identity map, or a map to a simple pitch set); (ii) a range of pitch set sizes from unisons to larger ones; (iii) symmetry of going flat and sharp; (iv) continuity in spirit, though not literally possible with discrete pitch classes. Idea (i) suggested a way of constructing such variants: looking for the more complicated pitch sets halfway between the simple pitch sets. If possible, such a halfway-set should be simpler than the trivial case (the chromatic aggregate). Idea (ii) can be restated: played vibrato becomes sounding ornamentation. (Extended glissandi are fun to play, but not that deep to listen to.)

I generally rendered the pitch sets as slowly rearticulated notes, without octave duplication of pitch classes. The articulation of notes, from pizzicato-like to slow washes, was driven by violin elevation since extreme articulations worked better as rare effects than as sustained timbres. An extra "filter" to thin out harmony was driven by bow distance. Only near the frog would all pitch classes sound, while nearer the tip fewer and fewer would sound until only unisons occurred.[151]

Intervals other than fifths are possible. Tritones generate too spare a harmony; semitone clusters, the opposite. Stacking fourths appears at first blush to work as well as stacking fifths, but it contradicts the sense of rising common to both playing sharp and moving to a sharper key. We do better with thirds. Major thirds stack to form an augmented triad; going from C to C# we would then have:

C – C E – C E G# – ... – C E G# F A C# – ... – F A C# – A C# – C#

or, when filled out so that at each step only one pitch class is added or removed:

C – C E – G# C E – G# A C E – G# A C E F – G# A C C# E F –
G# A C# E F – A C# E F – A C# F – A C# – C#

Minor thirds work similarly, building up diminished-seventh chords instead of augmented triads. Major seconds work if the game is played with whole tones instead of semitones (this also does not demand superhumanly accurate intonation). For example, a motion from C to D analogous to the C-to-C# above would be

C – C D – A# C D – A# C D E – G# A# C D E – G# A# C D E F# –

A# C D E F# – A# C D E – C D E – C D – D

From these examples we can see that using single intervals provides insufficient intervallic variety for all but the shortest compositions. But the rudiments presented here can be combined with each other and with other harmonic material to generate interesting vertical structures.

In this realm, notating exact sounding chords in a composition demands too much from the performer. Notating only the nearest tempered semitone with indications of bending the pitch sharp or flat, a little or a lot, leaves the exact harmonic richness up to the experience of the performer.

[151] This harmonic thinning could have been assigned to violin position, of course. It is assigned to bow distance because bow distance can change quite quickly, like harmony.

4.1.7 Matching controls and synthesizer

The violin is a set of strings just vibrating, but a common model of sound in synthesizers is individual notes with definite beginnings and ends. If we drive such a synthesizer with a violin controller, we must address the problem of deciding when a new note has started. (i) If during a single stroke the bow crosses strings, should the second string be a quick change of pitch from the old note, or rather a new note? If the latter, when does the old note end and the new one begin? This is nontrivial when the first string continues to resonate for some time, as is the case with an open string. (ii) If the bow changes direction without change of pitch, should a new note begin? (iii) If during a single bow stroke on one string a fingering change happens, should the old note continue with a pitch change or should a new note begin?

A modification of the fiddle~ algorithm improved its elision of notes for case (iii), even for a rising scale with fingers hammered down onto the fingerboard. I also tried using bow velocity as a hint for determining whether or not to elide two notes: if the bow has not changed direction, elision is indicated. Unfortunately motion tracking is slower than pitch tracking in the eviolin, so by the time the motion tracker reports that the bow has or has not changed direction, the pitch tracker has already decided if the note has or has not ended. Slowing down the pitch tracking to match the motion tracking would solve this, but of course worsen overall latency. It is also worth noting that with two trackers indirectly driving one parameter, a good design is for the slower tracker to store its value in a variable which the faster tracker (fiddle~) then reads and combines with its own data. Then the synthesizer is updated all at once, and subtle beating effects are avoided. If one parameter is updated at 30 Hz and another at 35 Hz, a mysterious phantom update can be heard at the 5 Hz difference frequency.

The universal answer to these three questions about note transitions is, of course, "it depends on the musical context." In other words, the performer needs to communicate this musical context to the instrument. A pedal acting like a damper pedal is one way: when the pedal is depressed, under all conditions the old note is never replaced by a new note but rather continued with a pitch change. But pedals are impractical for a perambulatory player, so instead of adding a secondary control I adjusted the innards of the pitch tracker until a fair compromise was reached in all three cases. I discovered that even the hardest case, an inaudible bow change, was possible by lifting the bow off the string for a moment and then sneaking it back onto the string while the string was still vibrating. Starting right at the bridge and then backing off from *sul ponticello* helps the bow's stick-slip motion resynchronize with the existing vibration, like easing up on the clutch of an automobile. (This is quite easy on the lower strings which have more inertia.)

Another possible matching of these two concepts is (with a separate signal from each string) having one "note" per string, always being played though possibly with zero amplitude. But the onset of a synthesized note may have some transient behavior at the attack, so a legato scale that crosses from one string to another would produce two attacks, not what the eviolinist intends. The concept of note seems to be closer to a single bow or a single musical breath than to a single string, but how this might be implemented is unclear. For now, from bowed-string control it seems best to drive only synthesizers without strong attack transients.

4.2 The eviolin in a virtual environment

I tried several ways of integrating the eviolin with a virtual environment (VE). The VE directed the performer by displaying a score, and also provided feedback to the performer to guide his actions.

The VE used is the CAVE at the University of Illinois National Center for Supercomputing Applications, a cubical room about 3 meters on a side. The operator wears LCD shutter glasses to properly see the stereoscopic images drawn on walls and floor by projection displays (figure 34). The CAVE tracks the position of these glasses, so the images are drawn from the operator's point of view with correct perspective. The net effect is that objects appear to be in the same space as the operator. One can walk around objects and inspect them from different angles, or manipulate them by means of a wand, a mouse-like object also tracked in position and orientation (Cruz-Neira, Sandin, and DeFanti 1993).[152]

In adapting the eviolin to work in the CAVE, I left out the SpacePad motion tracking because it would introduce extra complication with negligible benefit. The CAVE's own motion tracking, the Ascension Flock of Birds system, is far more accurate (and thirty times more expensive). It is needed anyways for the violinist's shutter glasses, and from the position and orientation of the violinist's head we can infer where the violin is. (The linkage between head and violin has little variation for any given violinist.) Not using the SpacePad means we avoid needing to transport, boot, and reconfigure the PC containing its circuitry (this PC normally acts as host for the eviolin). This convenience greatly simplifies training and rehearsal, not to mention development. The tether to the violin body is conveniently lighter, too, having only the cable carrying the audio signal.

The audio signal from the violin is sent via patchbays to an audio input of the multiprocessor Silicon Graphics Onyx computer which drives the CAVE. (Audio input has been used before in CAVEs, but only as off-the-shelf speech recognition software.) The same VSS software which analyzes the violin's audio signal and synthesizes sounds on the standalone PC also runs on the CAVE's Onyx. The audio output goes through speakers in the eight corners of the CAVE. In this early phase of research, moving the audio output around from speaker to speaker seemed premature; monophonic output has sufficed so far.

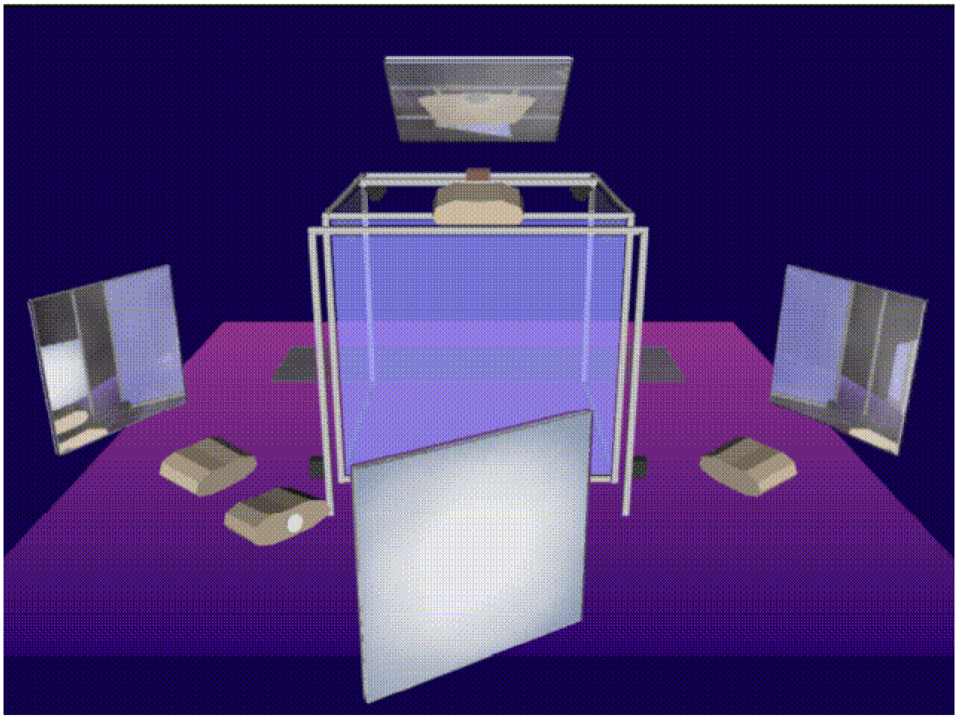


Figure 34. The CAVE Virtual Environment. Image by Milana Huang reproduced with permission from the Electronic Visualization Laboratory, University of Illinois at Chicago.

4.2.1 *Dusty Echobox Study*

An early VE I built, *Dusty Echobox Study*, is more an environment to be experienced than a composition with beginning, middle, and end. It grew out of the idea of standing in front of one of five microphones and modifying the sound picked up by each microphone in various ways. Since the microphones would be simulated, there was no need to emulate conventional cardioid pickup patterns or conventional microphone placement. The pickup pattern of each microphone is a cylinder running straight through the CAVE, 80 cm in diameter. These cylinders slowly tilt and move around, changing how the pickup patterns

[152] Of course one cannot *feel* these objects. One can walk right through them, since they are drawn on the walls. Cai et al. (1995) have built a two-handed force feedback system for the CAVE called Big SPIDAR. This system is accurate enough to simulate catching, dribbling, and throwing a basketball, having 14 mm resolution, 30 N force, 50 g inertia, and 2.5 msec latency. However, such techniques have not yet seen widespread use.

overlap. Each cylinder is drawn as a sunbeam illuminating dust motes of a particular color drifting through the space (figure 35). The sunbeam brightens as the output of that microphone's audio processing (filtered echoes) becomes louder. As the audio processing consists of filtered echoes, this is seen as pulsations in time with the echoes. The core of the sunbeam also brightens, the closer you come to its central axis. This VE, analytically simple but perceptually elaborate, blurs the boundaries of instrument, composition, and environment. It is an instrument, both acoustic and visual: there is immediacy of response to gestures, both spatial and violinistic. On the other hand, composition is the best word to describe the visual representation and the five different audio processing chains. Finally, environment describes better than instrument or composition how the five sunbeams move around independent of the performer's actions.

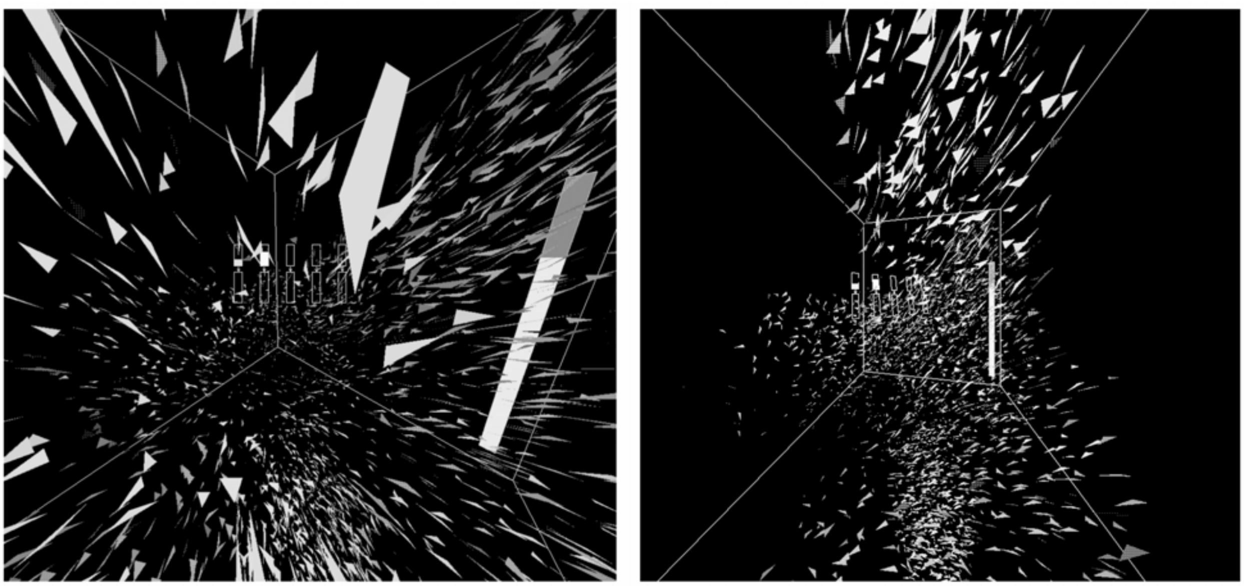


Figure 35. View of *Dusty Echobox Study*. The left image is from the eviolinist's point of view. The right image is the same scene viewed from outside the CAVE. The array of bar graphs indicates the amount of signal entering and leaving each delay line. The tall bar graph at right indicates the level of complexity of the environment, quite high at this moment.

This VE overwhelmed both violinists and audience. So much new material was presented at once that sensory bedazzlement blocked out any comprehension of verbal explanations of how it worked. (The situation is like an untrained person examining the cockpit of an airliner. Useful structure and organization becomes apparent only after much training in both what the displays mean and how to move one's gaze around them.) To build up in the eviolinist a mental model of the VE, the key was to strip out features one at a time until he felt in control of the situation, not merely reacting to unpredictable changes around him.[153] As preliminary training for both violinist and audience, I implemented a slider which continuously interpolated between the simplest environment and the full-featured one, so we could adjust the level of complexity from moment to moment in an adaptive training regimen. The simplest level has thin stationary vertical cylinders rendered as wire-frame models, not dust motes. As the slider value increases, the cylinders begin to move around slowly; then the cylinders tilt away from vertical to create a few overlaps between pickup patterns; then the wires fade out and dust motes fade in. Continuing, the cylinders widen to increase overlap between pickup patterns. At the highest complexity, the cylinders move around more rapidly. As a final expressive touch, echoes are entirely eliminated when playing at the tip of the bow.

4.2.2 *Urbana Nocturne*

Urbana Nocturne is more rigorously a composition, not an environment like *Dusty Echobox Study*. It has a fixed order of pitches with durations, in conventional staff notation. Other aspects of the music are rendered graphically in the VE. The life-size setting is the middle of a grassy field with three species of flowers, encircled by a ring of street lights 6 meters in diameter (figure 36). The sky is colored like sunset, darkening over the six-minute duration of the composition. The lights are the shape and color typical of older residential streets and parks in Urbana, Illinois.

The streetlights direct the eviolinist to walk to different places in the CAVE, causing the timbre of synthesized sound to vary. Normally they are all a dim yellow-gold color. When the eviolinist should walk to another place, the streetlights in that direction whiten while the ones in the opposite direction redden. In addition, the lights slightly pulsate in size and brightness, from the red end to the white end; the speed of the pulsation corresponds to the distance the eviolinist needs to travel. This redundancy ensures that the requested direction of motion is clear, no matter which direction the eviolinist is looking at. (While debugging this visual representation I also drew a bright spotlight on the floor, as well as a heads-up square map which can be seen in the middle of figure 36.)

[153] Part of this initial lack of mental model was the egocentric viewpoint of the violinist—they had to actively look around to find the other sunbeams behind them, to produce sounds associated with those sunbeams. But when the sunbeams were stationary, they could remember and associate a fixed spatial location with each sound.

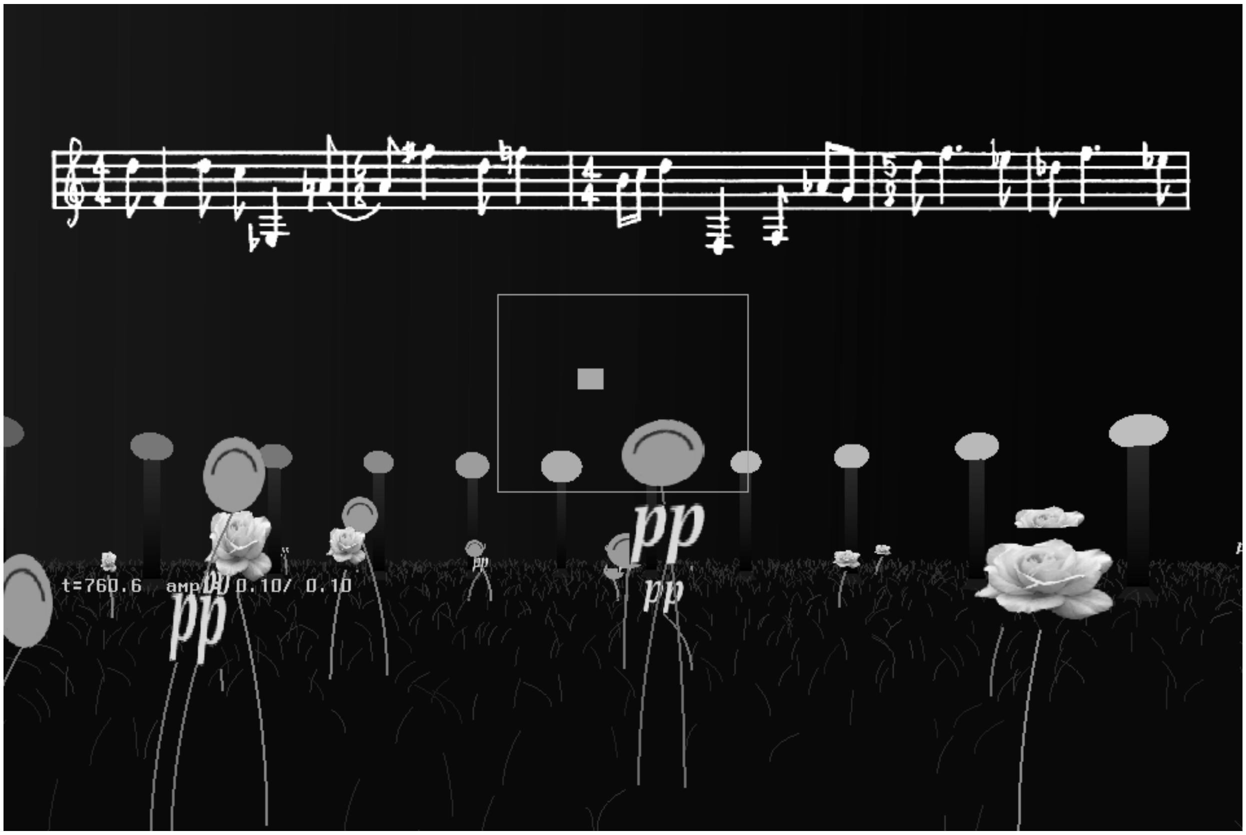


Figure 36. Image from *Urbana Nocturne*. The rose flowers are unmodified, indicating lack of "schmaltz"; the dynamics flowers indicate pianissimo; the disc flowers indicate legato; the staff indicates pitches and durations. The square at middle and text at left are diagnostics.

Each of the three species of flower is a linear family of texture maps with transparency (alpha channel in OpenGL terms), as is the set of one-staff score images. One species indicates dynamics: its flower-head ranges from literal *pp* to *ff*. The circular flowers contain glyphs ranging from an upward-pointing slur (legato) through a horizontal line (marcato) to a downward-pointing V (staccato with an attack). The rose changes from a natural pink to an extremely artificial blue, greater blueness indicating greater "schmaltzy" overexpressiveness. The one-line "pages" of the score turn to the next page automatically; a line growing from left to right paces the performer but without imposing detailed rhythm like the bouncing ball of Karaoke (or, for that matter, Max Fleischer's "Song Car-tune" films of the 1920's). The iconic directions given by the flowers change not instantly but gradually. Distant flowers change early, the change rippling through the field towards the performer until the nearest flowers change at the exact moment the performer should execute the change.

Since peripheral vision was important for noticing changes in the flowers in *Urbana Nocturne*, and peripheral vision was obscured by the shutter glasses, I decided to have the eviolinist wear the glasses above the forehead (like some wear sunglasses). This meant that the eviolinist saw images drawn on the walls, not in three dimensions. But this lack of stereopsis did not strongly reduce the immersive feeling of the environment, nor reduce the performance of the eviolinist, since except for the grass underfoot all objects were two to four meters away; stereopsis is much more important for distances under two meters. Had the distance been any greater, the images could have been displayed in fixed position, not adjusted to track the viewing position of the eviolinist. (The motion-tracked shutter glasses could not be entirely dispensed with, because their position was used to drive timbre.)

In both *Dusty Echobox Study* and *Urbana Nocturne*, playing the violin while walking around was a challenge in this CAVE. The only way a performer could avoid entanglement in the cables going to glasses, violin and bow was to meticulously remember the motions and turns they had taken, and retrace those steps in reverse. Eventually they would forget to do so, understandably so considering the other demands on their attention, and the cable snarls would begin.

4.2.3 Presenting musical scores in a VE

A large variety of activities and gestures are possible with a synthetic instrument in a VE. VE systems are typically in high demand, so performers cannot rehearse as much as they would like to. These facts combine to suggest that memorization of a composition is not practical unless many aspects of the music are improvised. So we should display a score in the VE. Notations on paper are of course possible, but they limit the performer's range of motion and/or obscure the view of the VE. Illuminating the paper may also interact adversely with the VE. Dispensing with paper, we should use the VE itself to display the score.

At this point the term score needs rethinking. A score consists of notations. It is a communication from composer to performer. It instructs the performer to execute actions with certain attributes in a certain order, at certain times if it explicitly represents time (as note heads or as a time line). In the terminology of control theory, performing from a score is precisely a multidimensional tracking task: play these pitches with these dynamics and these articulations, in this trajectory through time. Also, if the score has a fixed route through it (without mobile forms like Stockhausen's *Klavierstück XI* or Boulez's third piano sonata), the tracking task has infinite lookahead. (Performing *well* from a score is, of course, more than getting the notes right. We will return to this point in a moment, seeing how far control theory applies here.)

Certainly we can still display static notation-images in the VE, one “page” at a time if the whole score cannot fit at once. But the VE can also act as a conductor: not (just) one which beats time, but one which presents auxiliary instructions/reminders to the performer at or just before the moment they are needed. This splits the multidimensional tracking task into individual unidimensional tracking tasks. For instance, the VE may directly present instructions relating to dynamics (as a conductor beats a larger or smaller pattern from moment to moment). This changes tracking of dynamics from infinite lookahead to short or zero lookahead. This is a disadvantage, more so for some parameters than others (tracking pitches with zero lookahead is enormously harder). But the disadvantage is offset by reducing the amount of information presented in static score images. VE’s have limited spatial resolution of images; even high-resolution desktop displays show much less than a page of manuscript paper can. On the page, space is cheap but time is frozen; in a VE, space is expensive and time is cheap. So a profusion of dynamic markings on a page is appropriately replaced in a VE by a single indicator of dynamics changing in time. The general principle is to allocate the spatial resources of the VE proportionally to the need of each tracking task for lookahead, zero, short, or long.

A practical application of this principle taken in *Urbana Nocturne* is to present fragments of common music notation as only pitches and rhythms, while each of the other performance attributes is indicated by a sequence of time-varying indicators. Indicators farthest from the performer change earlier to provide lookahead. A disadvantage of this visual division is imprecise synchronization between the note-heads and the separated indicators: indicating large changes over the small temporal scope of a few short notes is difficult. Another problem is that the visual sense is poor for alarms since the viewer’s gaze may be elsewhere; the eviolinist already has much to look at, and easily misses when some of the flowers change from *ff* to *mp*. For temporally sensitive information, haptic display is a more reliable channel of communication than visual; since the visual channel has greater bandwidth, coupling the two would work well.

4.2.4 The VE as mediator between score and sound

Let us look again at the VE consisting of electric violin, CAVE, and software sound synthesizer. The VE presents instructions in various forms to the performer. It measures the performer’s response to these instructions—positions, orientations, pitch, loudness, and from these numbers synthesizes an acoustic and visual response. We go from numerically represented instructions through an imperfect performer back to a numerically presented performance. Could we not simplify this by going directly from the numerical score to its ideal performance? The actual performer is not necessary; a simulation of the performer is both simpler and more accurate.

At least that is how the control theorist thinks. In this paradox lies the difference alluded to above, the difference between performing and performing well. This unquantifiable adverb in no small part explains why we still have performers as well as composers of music. The more scholarly term is performance practice: those things not notated by the composer. And why not take advantage of a performer's intuition and experience? The extreme answer is: a desire to break completely with the past, or at least to limit its intrusion into one's own music. But leaving aside historical aesthetics for the moment, certainly *some* of a performer's training is good and useful, beyond mechanical tracking ability. To achieve a desired musical result, it is simply more efficient to divide the effort between performer and composer than to attempt to notate everything. (Some zealous editors of the Beethoven piano sonatas have demonstrated what can happen when the performer is too little trusted.[154])

So the elaborate apparatus of presenting instructions to the performer and measuring the resulting performance does have a purpose. The performer interprets these instructions, perhaps makes errors of tracking (recall William Heiles's opinion of note-perfect Beethoven), and also adds something to the data, something resistant to quantification and analysis because such would be tantamount to quantifying and analyzing the performer's entire musical experience.

We can draw a corollary from this: it is acceptable for the VE to measure aspects of the performer's behavior about which it does not give the performer instructions. This is acceptable but not necessary: the performer may find enough room to play in, simply in the precise specification of approximate values (fuzzy discrete rendering) and the precise synchronization of multiple parameter streams. But we cannot accuse the VE of spuriously extrapolating data if it measures something which the composer has not specified. Even measuring something which the performer is not able to control attentively during performance can be musically justified, if the performer has enough rehearsal time to develop patterned responses.

We draw another corollary from this mode of interaction between VE and performer. Since the point is not "note-perfect" tracking of instructions, the VE can be thought of as giving suggestions, not commandments. In our Beethoven piano sonata example, wide interpretation of these suggestions leads to controversial performances (re-compositions?) such as those of Glenn Gould. But even Beethoven's piano sonatas only suggest some things: exact tempo; dynamics such as the impossible crescendo on a pianissimo chord at the end of the first movement of Op. 81a, "Les Adieux"; and rate of decay of string vibrations, critical to some pedaled *recitativo* passages. Some of the instructions from the VE can

[154] All this implicitly assumes a well-trained, competent performer. Someone unfamiliar with Classical performance practice will do poorly with a Beethoven *Urtext*. But, to be fair, we also demand training and competence from the composer.

therefore also be suggestions: not to be whimsically ignored, but to be followed as much (and at such time) as makes sense to the performer, in the context of everything else happening at that point.

Applying this principle, the streetlights in *Urbana Nocturne* only suggest where the violinist should stand in the CAVE. This could have been a clear command, for instance by drawing a bright target on the floor; instead, the deliberately indirect pulsing streetlights guide the eviolinist to where the score says he "should" stand. It is up to the eviolinist to gradually decide in rehearsal how much attention to pay to the streetlights, at each moment of the composition.

Finally, since considerable effort goes into constructing the VE it is often desirable to show the VE to the public as part of the performance. This is particularly so if the VE is visually compelling, not just a collection of bar graphs and technical notation. If this is done, graphic designers should beware of the inherent compromises of a display intended for two very differently trained individuals, the performer and the audience member.

4.3 The impact of real-time performance

Composing for a human instrumentalist rather than inhuman tape playback raises certain issues. The two have different properties, to put it mildly. The instrumentalist as opposed to the tape machine has finite timbral range, finite accuracy, limited attention, the many social connotations of simply being a human with a human audience, and (most fundamentally) the ability to make decisions.

Babbitt notes that a recording of a Tchaikovsky symphony is a work of electronic music. The recording is a notation which specifies many things which Tchaikovsky chose to leave implicit, particularly the moment-to-moment variation of tempo. The symphony itself, like most compositions with human performers, has in practice not precise values but only ranges of tempo and approximate accelerations. Precise values are left up to the ensemble, depending on facility, lung power, room acoustics, and so on. Vaguely, this is an advantage over a performance realized on tape in its "liveness," its adaptability to the situation of each performance. Its disadvantage is its unreliability, its dependence on performers who can choose a good tempo and stay synchronized with each other. The trade-off is between security and adaptability.

Tempo is by no means the only parameter for which human performers inevitably produce ranges of values, not precise values (hence the introduction of the term fuzzy discrete rendering in the first chapter). Amplitude and certain timbral parameters particularly come to mind. Here the amount of precision and the space for the piece to "breathe", to vary from one performance to the other, is *contributed* by the performer, coming from his experience as performer.[155] The high accuracy of tape playback is not so important with these parameters, since our perception of them is inaccurate anyways. (This is circular reasoning, of course. If we could hear them accurately, so could the performer. Then by listening he could accurately play them.)

In a concert performance, besides composer and performer there are listeners. The task of the listener is easier with a live performance than with a tape performance because visual correlatives to the sound stream help him follow the structure and direct his attention. (Of course a tape piece could have an accompanying visualization of a synthetic performer-with-instrument. Kaper and Tipei (1998) have investigated this with the M4CAVE system; Pellegrino (2000) lists many amateur music visualization projects.) Sympathetic kinesthesia with a human performer can also enrich a piece.

Rehearsal develops and improves the piece as the performer sees it from different angles at different stages of learning. At a basic mental-model level, this is reinforced if the performer undergoes *part-task training*, guided exercises which emphasize certain techniques while omitting others, to learn aspects of playing the instrument.

In a real-time performance, the place where certain decisions are made moves from the composer's desk to the performer's hands. The composer has the leisure to contemplate all aspects of a decision, but the performer can handle only so much at once.[156] (One might facilely say that Ferneyhough denies this, leaving no room for the performer to decide anything.) So given the instrument and a performer of finite skill, the composer does well to consciously note these decisions and divide them appropriately between himself and the performer(s). In other words, the finite skill of a performer on this instrument determines what should be notated, and down to what level of detail.

[155] Recall also from the first chapter the difficult music of Ferneyhough where performers contribute their own particular ways of falling short of perfection.

[156] But composers with deadlines can sympathize with the overtaxed performer. They too cannot consider every ramification of their decisions before the score is delivered.

5. The Eviolin in Chamber Music

Timbre, in the general sense of varied instrumental combinations, is the primary organizing parameter in the chamber piece *COALS*. In the first movement, timbre defines large-scale structure while a serialized family of pitch/rhythm motifs defines middle-scale structure. In the second movement timbre is serialized (though in a sense closer to Xenakis than to Schoenberg) together with pitch material and dynamics. This total-serial organization is applied only at the smallest scale, imposing short sequences on a large number of units each only several seconds long. These sequences gradually merge to produce a global structure true to the base materials. *COALS* distils many concepts from chapter one (Boulez's generalized serialism, Xenakis's stochastics, Babbitt's algebraic structures, Wolpe's freedom) into a single artistic whole, using a synthetic instrument made possible from the heuristics of instrument design outlined in chapter two and the simplicial interpolator presented in chapter three.

5.1 Compositional criteria

COALS is a large-scale composition for eviolin and five orchestral instruments. It demonstrates the eviolin and methodically explores its possible relationships with other instruments. A large universe of possibilities is severely constrained to shrink it to a manageable size, then thoroughly mined. This approach, maximum variety from minimum materials, owes a debt to Babbitt (or perhaps to the Vulcan proverb, "infinite diversity in infinite combinations"). Deliberate manipulation of structural parameters in the compositional process produces interesting patterns of growth and contrast, ones which my instinct alone would have been insufficiently imaginative to discover. A side effect of the methodical exploration of sub-ensembles in *COALS* is that the piece feels light on its feet, without long tutti sections (a brief ten-note tutti occurs in bar 488, for example). This increases the perceptual density of information; paradoxically so, since the same work with some of these longer rests filled in would actually contain more data, but be at greater peril of being perceived as "everyone's playing all the time." More formally, this constant flux of sub-ensembles stays fresh even over a long time span by avoiding predictability.

The "broken consort" sound is inevitable with a newly invented instrument, unless we have the funds to build a whole family of such instruments. So *COALS* uses a heterogeneous ensemble. Having accepted heterogeneity, the other instruments may as well come from different families. A representative combination of families, monophonic and polyphonic, percussive and sustained (and both: plucked and bowed strings), is given by flute, clarinet, eviolin, violoncello, trombone, and electric bass guitar. Since some instruments in this ensemble are based on conventional equal-tempered semitones, and since all produce harmonic spectra, the sound synthesizer of the eviolin therefore also produces pitched sounds with harmonic spectra so the eviolin can work as a unit with the other instruments.

The physical gestures of the eviolinist are fairly quickly understood by listeners in how they affect the sound produced, so the score needs no specifically didactic elements which introduce one parameter at a time in clearly isolated contexts.

In terms of the five sonic dimensions postulated by Boulez—pitch, duration, amplitude, timbre, spatial position—the eviolin is not as different from the other instruments in the ensemble as one might anticipate. This comes from wanting not to make *COALS* a concerto for eviolin, but to arrange the instruments as structural equals. Certainly the eviolin attracts attention in performance because of its novelty and the spatial motion of its performer, but this need not be emphasized. The electronic synthesis of the eviolin could let it play far louder or quieter, more extremely and more subtly than the other instruments; recall the discussion of the aesthetics of listener saturation in chapter one. But this is not needed; if anything, after considering this aesthetic of "more" I find that it leads even more inexorably to the end of fertile lands than total serialism did. (But despite all the structures which enforce *égalité* and *fraternité*, *liberté* often favors the eviolin for solo roles.)

In particular, the eviolin's pitch range, duration of short and long notes, and dynamic range are all similar to those of the other instruments. In each of these aspects, fineness of resolution is also conventional. Spatial position is fixed, for better blending with the ensemble; if only one instrument could cast its sound around the hall, solo/tutti opposition would be inevitable. But in timbre the eviolin outstrips its nonsynthetic companions. The range of spectral richness or thinness, brightness or dullness, produces a wider range of distinguishable timbres than is possible from the other instruments. From these stipulations, the structures which can be set up in each of Boulez's five sonic dimensions will therefore be similar for all the instruments. Only in timbre does the eviolin go farther; in particular, it has two independent controls of timbre. This allows for quasi-modular structure: a gesture along spectral brightness (say, dull to bright and then back to medium) can be repeated with different values of spectral richness. This is of course not strictly modular in the sense of subdividing a continuum, but does produce structures which are heard similarly to pitch motifs repeated at different octave registers.

5.2 Structural analysis of COALS

COALS is in two movements. The first one, *Camera obscura*, is about 7 minutes long. The second, *Carta terrae incognitae*, takes 15 minutes. For brevity we often refer to the movements simply as *Camera* and *Carta*. (It is easy to remember which is first, from their alphabetical order.) The movements share some formal structures; certainly they share the same instrumentation. We first consider the things they have in common and then go on to particularities.

Inevitably this analysis is incomplete, almost exclusively describing the systematic aspects. The aspects which are imaginative (indisciplined, to use Boulez's word again) are better heard in the music than read in an analysis.

5.2.1 Timbre

The gamut of raw timbral resources is mostly continuous, not discretely divided. Pizzicato and harmonics are the only discontinuous aspects of timbre, but even these are made to be continuous at a few points. In bar 115 of *Carta*, the cello bow smoothly picks up a decaying pizzicato note. Pitches repeated or continued with various spellings of natural harmonics occur regularly in the eviolin part: bar 61 of *Camera*, and bars 168, 354 and 525–528 of *Carta*. (The cello also does such in *Carta*, bar 352.)

A lattice of modules of timbre, the multidimensional timbre offered by a collection of diverse instruments, was not adopted for *COALS.* Though intriguing at first, the lattice's considerable structure seemed unwarranted. Instead continuous and discontinuous paths are traced through a timbre space built up of instrumental combinations, detailed below in connection with figure 43.

As a point of departure for grouping instruments, *COALS* uses the simple model provided by Babbitt's *Composition for Four Instruments*, namely using each combination exactly once. But within this egalitarian model still more criteria for grouping apply. Pairings exist within this ensemble, for example bowed strings, woodwinds, plucked strings, electrically amplified instruments. But a chain of pairs is not possible as with the ensemble of *Le marteau sans maître*.[157] Instead we generalize the linear chain to a more interesting topology, one of relative distance determined by tessitura, frettedness, ability to play several pitches at once, and so on (figure 37). (This is literally a topology: in figure 37, {{f, c}, {t}, {e, v}, {e, b}, {e}} is a basis for a topological space whose open sets (unions of these basis elements) are the sub-ensembles whose members share some of the attributes listed in the figure.) So depending on the view one takes, this heterogeneous ensemble has both unity and variety of timbre.

[157] Recall these pairings from chapter one: contralto voice and alto flute share human breath, alto flute and bowed viola monody, plucked viola and guitar the plucked-string timbre, guitar and vibraphone long resonance, damped vibraphone and xylophone pitched mallet timbre.

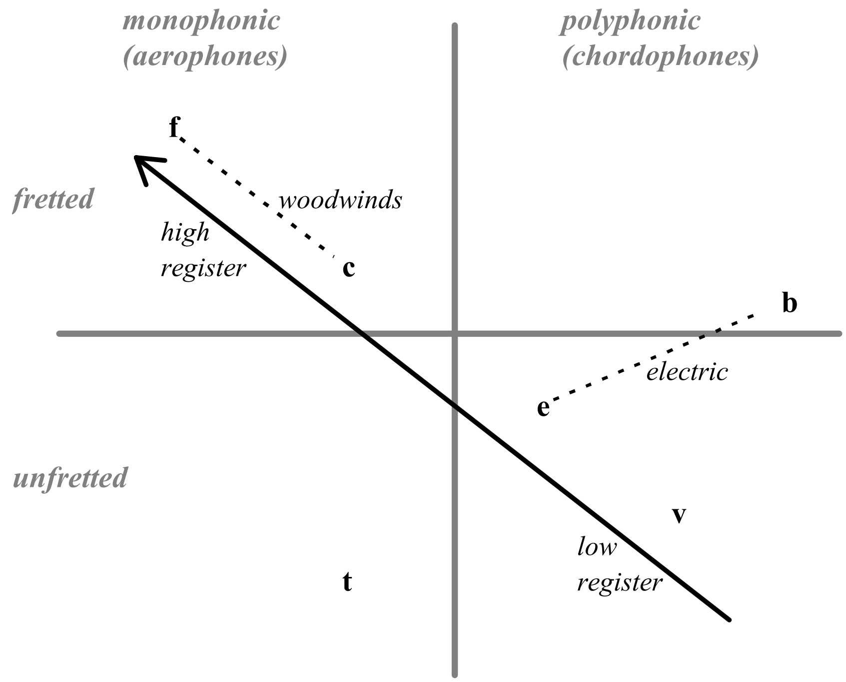


Figure 37. Various groupings of the instruments used in *COALS*, arranged to suggest relative distance between instruments. Instruments are abbreviated: (f)lute, (c)larinet, (e)violin, (v)ioloncello, (t)rombone, (b)ass guitar.

These pairings induce a coarse measure of timbral distance between instruments seen in the planar distribution in figure 37. These pairwise distances between instruments are then used to specify distances between sets of instruments. In *Camera* this set-distance emphasizes sub-ensembles with strong internal similarity (*e.g.*, flute, clarinet, and trombone) or strong internal contrast (*e.g.*, cello and flute). We will see that in *Carta* set-distance exhaustively defines a timbre space (figure 43), where the emphasized sub-ensembles are the ones more timbrally distant from others (*e.g.*, flute and bass guitar at lower left; flute, clarinet, and eviolin at bottom center).

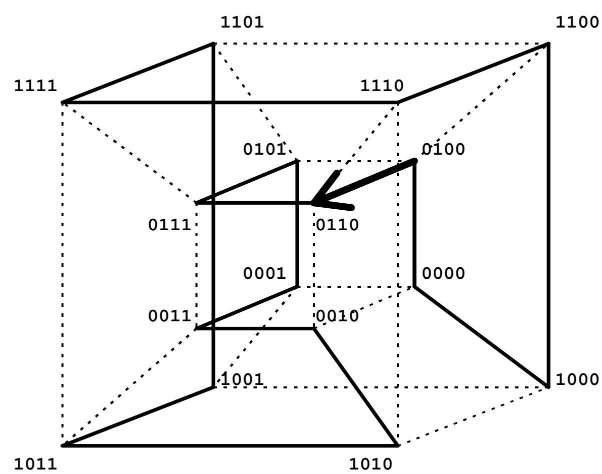
5.2.1.1 Timbre in *Camera obscura*

A germinal idea for *Camera obscura* was a traversal of all subsets of a set of size 4, *i.e.*, an ordering or serialization. This ordering is listed in figure 38a. Successive elements of this ordering differ by the addition or removal of exactly one element (exactly one digit changes between one and zero), making it what engineers call a Gray code. This particular code has two more properties which suit it to musical interpretation. First, it has a temporal asymmetry: compared to most Gray codes on four elements it has a wide spectrum, *i.e.*, a variety of lengths of presence and absence of elements (the length of a run of ones or zeros in figure 38a). Second, it is spatially approximately symmetric in that each of the four elements has about the same number of presence/absence transitions. This can be visually verified: no single axis predominates in figure 38b.

An elaborated version of the code in figure 38, a Gray code on six elements, is the starting point for defining the succession of macroscopic timbres in *Camera obscura*, namely the $2^6-1 = 63$ nonempty sub-ensembles of six instruments. The literal succession defined by this code is split into five parts. The actual succession used to construct the score comes from shuffling the parts to form a pleasing overall shape. The boundaries between these five parts are carefully chosen to cause large discontinuities when the parts are alternatively juxtaposed as a result of the shuffling. Discontinuity here means the sudden entry or disappearance of considerably more than one instrument, such as the transition from solo clarinet to full ensemble at measure 131.

```
→
0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0
1 1 1 1 0 0 0 0 0 0 1 1 1 1 0 0
0 1 1 0 0 1 1 1 1 0 0 1 1 0 0 0
0 0 1 1 1 1 0 0 1 1 1 1 0 0 0 0
```

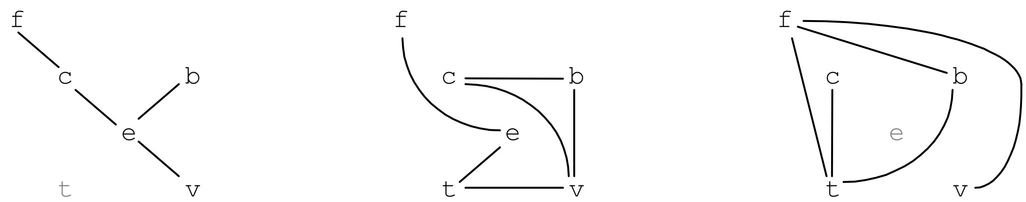
(a)



(b)

Figure 38. (a) A special Gray code on 4 elements. (b) Graphical derivation of this code. Vertices linked by an edge have three digits in common. A one in the first, second, third or fourth place corresponds to a point on the outside, top, front, or left half respectively.

In *Camera* the duration of each instrumental sub-ensemble comes from the internal similarities and dissimilarities of that sub-ensemble. In figure 37 we saw the instrumental groupings underlying the distances between instruments. Figure 39 extends this to a coarse measure of distance between instruments: pairs of instruments are considered similar, neutral, or contrasting depending on how many attributes are shared or opposed in the pair.



(a) (b) (c)

Figure 39. Relative distance between instruments, based on the pairings shown in figure 37. Lines indicate (a) similarity, (b) neutrality, and (c) contrast. Instruments are abbreviated as in figure 37: (f)lute, (c)larinet, (e)violin, (v)ioloncello, (t)rombone, (b)ass guitar.

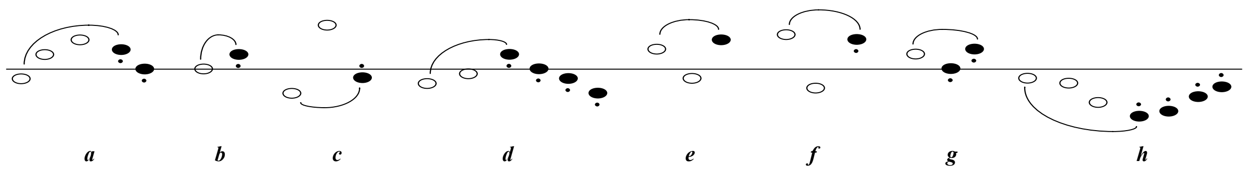
Within a sub-ensemble, then, several pairwise comparisons are possible. For brevity we here consider only the case of three instruments, in which three pairwise comparisons occur; similar patterns hold for sub-ensembles of other sizes. Figure 40 shows how sub-ensembles with strong similarities and/or contrasts receive longer durations. These durations are related exponentially, not linearly, since this is how we perceive duration (recall the critique in chapter one of Xenakis's vector spaces). These raw duration-ratios then slowly vary on the scale of the whole movement, to create a variety of harmonic rhythms.

Similar pairs	**3**	**2**	**2**	**1**	**1**	0	0	0	**1**	0
Neutral pairs	0	0	**1**	**2**	**1**	**3**	**2**	**1**	0	0
Contrasting pairs	0	**1**	0	0	**1**	0	**1**	**2**	**2**	**3**
Relative duration	**8**	**4**	**1**	**1**	**2**	**1**	**1**	**2**	**4**	**8**

Figure 40. Relative duration as a function of the number of similar, neutral, and contrasting pairs of instruments in a sub-ensemble of size 3. (Since three pairs exist in such an ensemble, the top three numbers in each column sum to three.)

5.2.1.2 Pitch, gesture, and rhythm in *Camera obscura*

We have seen the structural outline of *Camera obscura* at the level of timbre. Within this there are gestural and pitch structures. Figure 41a lists the eight primitive gestures from which the movement is built up.



(a)

D E F Bb B	B D	A C Db	D# E F Gb Ab Bb	E F Bb	F# G Ab	B D Eb	C Db Eb F# G Ab A
a	***b***	***c***	***d***	***e***	***f***	***g***	***h***

(b)

D Eb E A Bb	Eb D	C Db F	E F# G G# A Bb	E A Bb	Gb Ab B	D Eb G	C Db F F# G Ab B
a	***b***	***c***	***d***	***e***	***f***	***g***	***h***

©

Figure 41. (a) Eight gestures of pitch contour. (b) Eight corresponding pitch class sets. (c) A family of alternate pitch class sets.

These structures of pitch and gesture are tightly correlated. Each of the gestures *a* through *h* is rendered exclusively with its corresponding pitch class set in figure 41b. For example, at the very beginning of the movement flute and clarinet clearly state gesture *a* with pitch classes (in order) B, E, F, D, Bb. Since these gestures have considerable freedom of interpretation in the musical surface—dynamics, assignment to instruments, durations, and so on—this is not as restrictive as it first seems. Some of Beethoven's motivic repetition is far more constrained than this, for instance. The sequential order of pitch classes in each gesture is free as well, as long as the relative size of intervals in the gesture is respected.

At bar 59 eviolin and cello introduce a derived collection of pitch class sets, stating gesture *a* on the pitches Bb D Eb E A. This derived collection comes from remapping individual pitch classes: the whole-tone scale on C remains unchanged, but the whole-tone scale on C# is transposed up four semitones. This results in pitch sets which have a recognizable overlap with the originals, yet also have different internal structure because semitone intervals become four- or five-semitone intervals (figure 41c). This new pitch mapping remains in force until a solo clarinet passage builds up into a *tutti* return at bar 131 of the original pitch mapping.

Figure 42a restates figure 41b to emphasize the common elements among these pitch sets. To produce a slow circulation through the chromatic aggregate, the gestures/pitch sets are serialized in cyclic fashion as shown in figure 42b: *abe*, *bcg*, *cfh*, *bfg*, *bdeg*, *beg* go completely around a circle in clockwise direction.

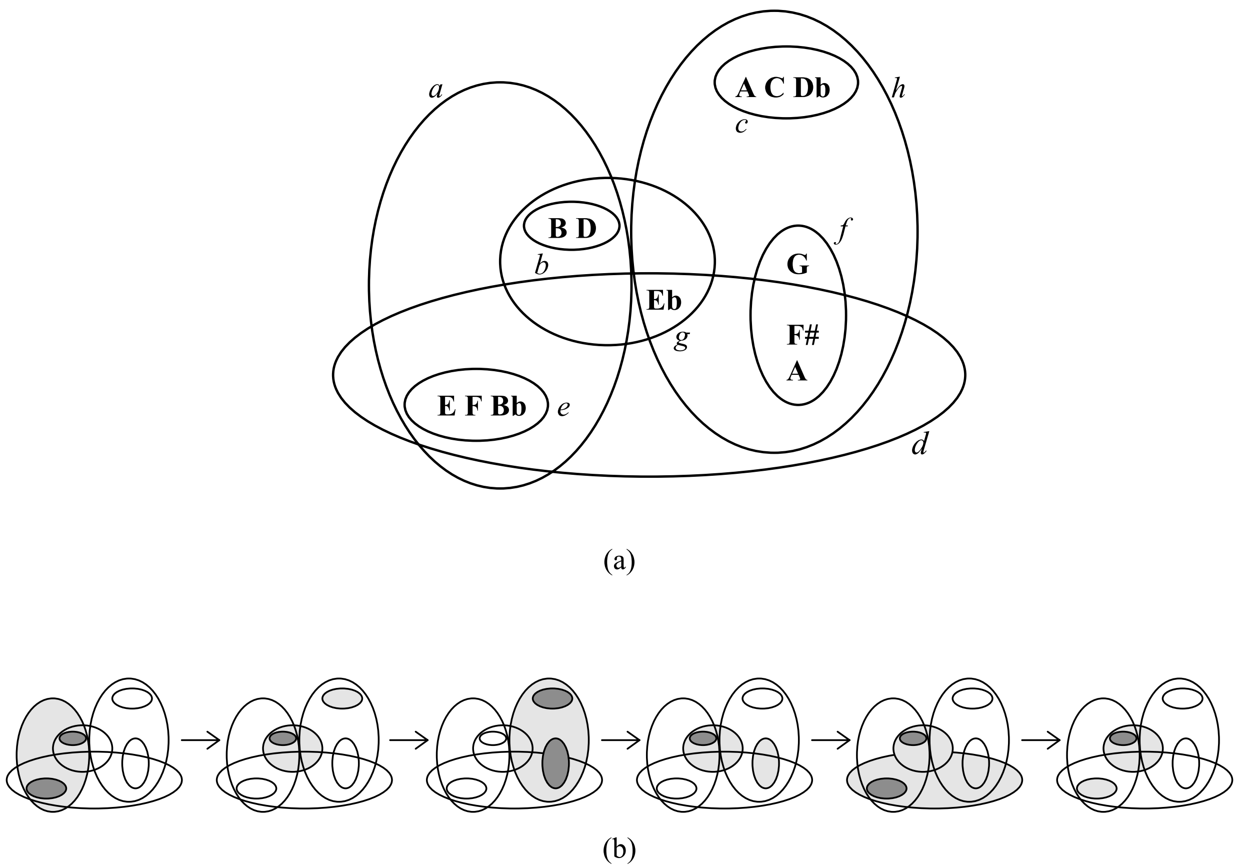


Figure 42. (a) Venn diagram showing common pitch classes in pitch sets *a* through *h*.
(b) Serialization of these pitch sets and their corresponding gestures.

As this circulation proceeds, the most common gestures/pitch sets are *b* and *g*; looking within pitch sets, the most common pitch classes are B, D, and Eb. These tend to be de-emphasized in the musical surface by appropriate choice of dynamics, register, and instrumentation in order to increase the overall information density of the movement. On the other hand, the less common gestures/pitch sets *a*, *d*, and *h*, together with pitch classes C, C#, A, and G, are generally reserved for rendering more dramatic structural events. This correlation of rare things with rare things actually decreases information density but clarifies structure: several new things occurring simultaneously are more easily recognized as a structural articulation.

The circulation is sometimes reversed and sometimes initiated at a different place, corresponding in traditional serialism to retrogression and Stravinskian rotation respectively. The speed of circulation also changes as the movement unfolds, in order to vary the harmonic/gestural rhythm.

In summary, there are correlations between pitch, rhythm, and gestural content in *Camera obscura*. These are analogous to the correlations between timbre, dynamics, and pitch content in *Carta* (figure 43), to which we now turn.

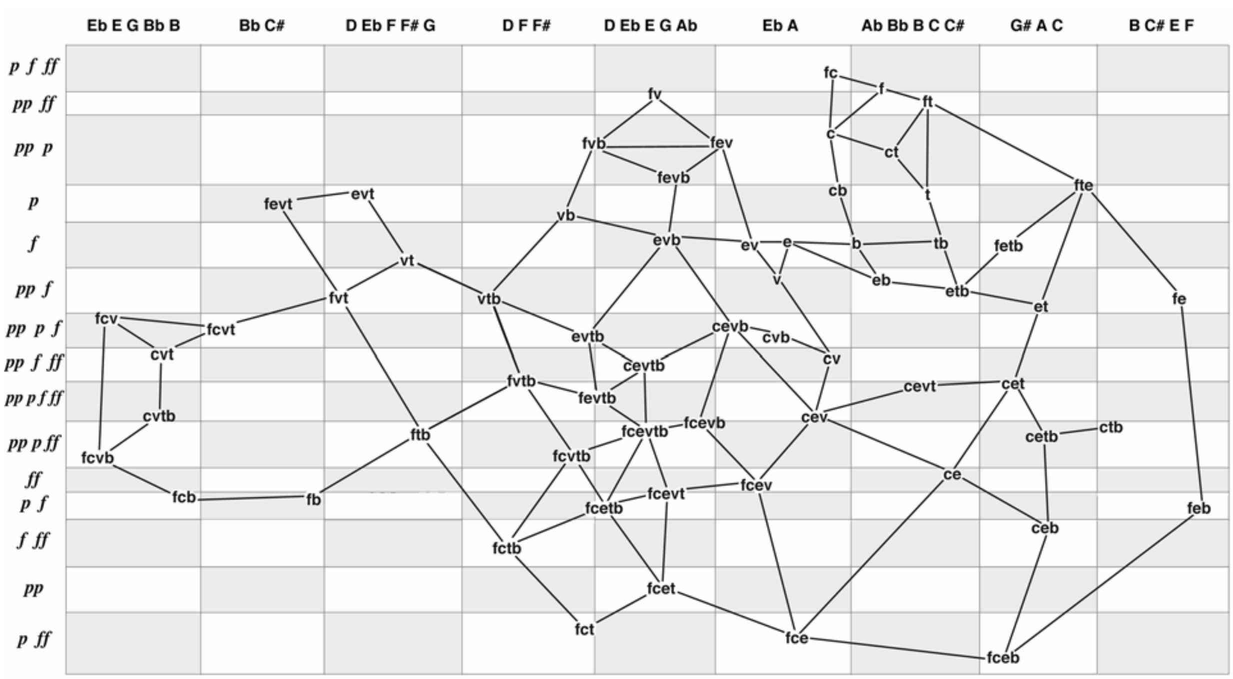


Figure 43. The space of timbres in *Carta terrae incognitae*, the distance between all subsets of instruments. Instruments are abbreviated as in figure 37: (f)lute, (c)larinet, (e)violin, (v)ioloncello, (t)rombone, (b)ass guitar.

5.2.1.3 Timbral definition of structure in *Carta terrae incognitae*

There are $2^6-1 = 63$ points in the space of timbres shown in figure 43, namely each nonempty subset of the whole ensemble. The six instruments are not equidistant in figure 37, so the space is naturally represented by a 6-cube distorted by these unequal distances.[158] The constraint imposed on the space represented in figure 43, then, can be imagined as a crinkled planar slice or shadow of this 6-cube. This crinkling produces the irregular mesh in figure 43; it further distorts the distances, but reduces the number of edges in the graph.[159] Though the planar graph suggests a triangulation of this set of points, it is not complete in this sense: the thinned-out triangulation is also an effort to reduce the number of edges. The reason behind this, in turn, is to keep the structures built on this graph down to a manageable size.

The edge set of this graph is partitioned with directed paths. The paths are constructed iteratively by depth-first traversal of the edges. Each path begins at a vertex chosen uniformly randomly; whenever a path reaches a vertex with more than one edge leaving it, an edge is chosen also uniformly randomly. Each edge is used only once. When a path gets stuck because no more edges leave the vertex it has reached, it is called complete and a new path is begun. This process continues until all edges have been used.

This partitioning is actually done twice. This creates redundancy in the paths and in the musical surface creates recognizable repetitions of timbral successions, rather like slow chord progressions. These repetitions are not very long, though, since for any given edge of the graph the two paths containing it, one from each partitioning, will diverge from each other not far before or after that edge. The second partitioning also adds a constraint to the depth-first traversal: path lengths are bounded. This produces more shorter paths than the first partitioning and thereby increases variety.

The paths are arranged in small sequences to imaginatively balance conjunct and disjunct motion in all three parameters, timbre, dynamics set, and pitch set. A concatenation of a few long paths is continuous in timbre and discontinuous in dynamics and pitch, while a concatenation of many short paths is discontinuous in timbre (at the boundaries between paths) but can be continuous in dynamics and/or pitch. An example of the former is found in the slow orchestral change but rapid harmonic rhythm of bars 445–467. The latter is exemplified by bars 285–298, an almost Mahlerian fluctuation of instrumental forces supporting slowly evolving harmony and a gradual curve from *f* to *pp* and back. Within each path, the duration of an edge simply corresponds to its length in figure 43.

---

[158] A formal definition of distance between sets of instruments, derived from the distances between individual instruments, seems too laborious when only six instruments are involved. Such a set is still tractable for manual computation. For larger sets the formal definition would likely include ideas such as the Levenshtein distance (“minimum editing distance”) between two strings (Levenshtein 1966, Allison 1984). Mean pairwise distance, though simple, is inaccurate for cases like the distance between the set of all six elements and a subset with one element removed.

[159] This distortion extends to a structural level the “local indiscipline” favored by Boulez. His words bear repeating: “at the overall level there is discipline and control, at the local level there is an element of indiscipline—a freedom to choose, to decide and to reject” (Boulez 1976, 66).

The temporal model has one other level: these small sequences are themselves sequenced imaginatively to create the overall form. The largest-scale form thus gradually anneals or emerges from the low-level material itself, not from an arbitrary external model. It is a cooperation between formal structure and imagination. Babbitt's question of what makes something appropriate to map into a musical parameter is easy to answer here: large-scale ordering comes not from a skyline or coastline but from the innards of what has been already composed. No charge of inappropriateness can even be laid.

Global intensity contours are then overlaid on this largest-scale sequence. In the musical surface, overall intensity is rendered as intensities of individual instruments. Intensity and dynamics are entirely independent parameters: an intense *pp* and a low-intensity *ff* are just as possible as a high-intensity *ff*. Flute and clarinet intensity is defined simply as register, since these instruments produce such different timbres—warm, thin, penetrating—when lower or higher notes are played. Among its other controls, the eviolin's intensity control is forwards/backwards position, which drives spectral brightness: the backward extreme is nearly sinusoidal, and the forward extreme approximates a buzzy square wave. Cello intensity is rendered as sul tasto / sul ponticello; pizzicato is sometimes used as an extreme case of sul pont. playing. The trombone uses both a plunger mute and bell position; these could be used independently but *COALS* uses them as a unit. Low to high intensity therefore ranges from muted through *modo ordinario* to bell-up unmuted. The electric bass guitar uses a pair of pickups for intensity control: low to high intensity involves selecting the bass pickup, both bass and treble, or treble only. The player does this by operating switches or knobs on the guitar body. No other effects units are used with the guitar, though its volume knob sometimes sustains or increases amplitude during a note.

The rendering of a global intensity contour as a collection of particular intensities of individual instruments is comparable to how a pitch set or dynamics set is rendered as a set of simultaneous pitches or dynamics of individual instruments. These three rendering techniques, intensity, pitch, and dynamics, share yet another property: sometimes one instrument takes on several values at the same time, for example when more values need rendering than there are instruments playing. As pitch this is commonly seen as a trill; in dynamics an instrument jumps between several levels in a phrase or sequence of little separated neumes; in intensity, jumping between levels is again the case.

In summary, these formal structures agrees with the spirit of total serialism in that they serialize many musical parameters together with pitch. But the structures here are richer—and deliberately more perceptible—than the literal mappings of *Structures Ia*.

5.2.2 Large-scale temporal organization

Motives, gestures, and phrases vary in size in *COALS*. Some are only a bar long, others are longer. There is a spectrum of motive-length and phrase-length, like the range of phrase lengths in the music of Mozart. But unlike the music of Mozart, phrases in *COALS* do not form hierarchies. A better word may be parallelarchy. In everyday life this idea is seen more often spatially than temporally, for instance weather maps which indicate the variations of precipitation, temperature, barometric pressure, and so on over some region of the landscape, that is, a spatial rather than temporal extent. These properties correlate to various degrees. Where all these properties simultaneously take on extreme values or rapidly change in value, we apply dramatic names like "cold front" or "tornado." But this is the exception; both in weather patterns and in *COALS*, particularly at smaller levels, phrases tend to overlap rather than line up. Not everybody arrives at once.

At the time scale of a minute or so, sections are elided rather than simply concatenated. These joints between sections are heard differently by different listeners, depending on which attributes they attend to more in figure 43: timbre, dynamic content, or pitch material (or intensity, though this is a little more abstract than the first three). Each rule for sectional division suggests a particular reading of the musical structure, as shown for an excerpt in figure 44.

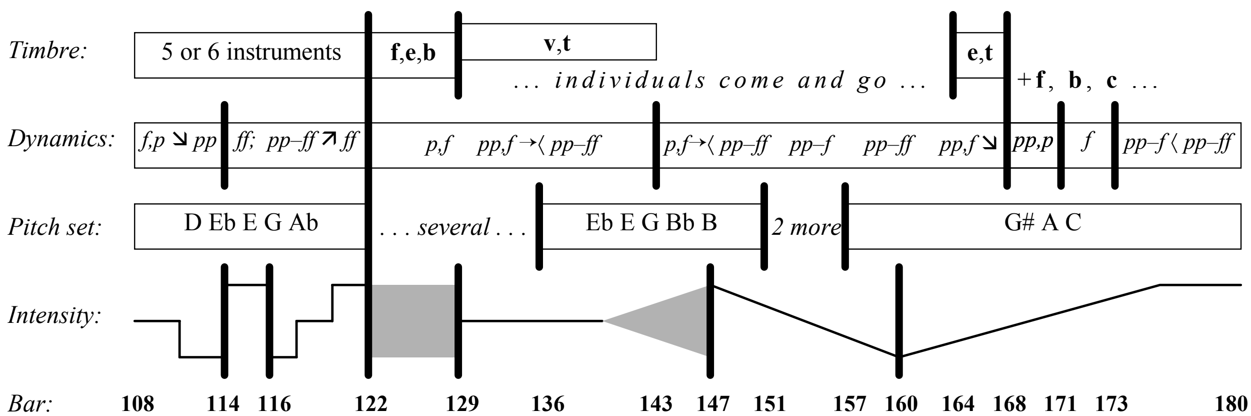


Figure 44. Different parsings of sectional division in *Carta terrae incognitae*, bars 108–180.

COALS fluidly moves between polyphony and heterophony instead of building up nested aggregations, the polyphonies *of* heterophonies *of* polyphonies suggested in (Boulez 1963a, 119–141). In particular, one class of heterophonies which becomes more and more prominent from bar 314 to the end is the duet. In this context a duet is any grouping of two instruments, possibly with other instruments playing simultaneously. For example, in bars 314–317 cello and bass guitar engage in a loose exchange of quiet mutterings; the cello then pairs with eviolin to bar 323 in a lyrical ascent while the bass protests in brief outbursts. As the movement progresses, the two parts of a duet become less independent. In the very last duets such as the unaccompanied clarinet and eviolin in bars 535–540, heterophony even begins to allude to unison playing.

5.2.3 Tempo and duration

Pulse occasionally sets up a regular, striated time amenable to acceleration/deceleration. Particularly *Carta* has occasional pulse trains for a dozen or so beats. For example, bars 92 to 94 thereof decelerate through a running eighth note passage in eviolin. (This gradual deceleration and sudden *a tempo* are strengthened by exact correlation with amount of flute vibrato.) In a related passage in bars 125–128, a "walking bass" line suddenly accelerates in mid-stride. Where tempo changes occur, both gradual and sudden (or momentary as with a fermata), the exact amount of change is left indeterminate. But for the majority of both movements the musical surface has no pulse, creating irregular, amorphous time with no possibility of acceleration or deceleration but only a variable density of events. No simultaneous multiple tempos therefore need occur; a single conductor beating a single pattern suffices to synchronize the players.

Highly flexible use of grace notes makes pulse by and large inaudible despite the quarter-note tactus and simple time signatures often evident on the page. This is particularly the case with longer runs of grace notes, which by the standards of common practice cannot fit into the time provided. It is up to each performer to decide where to steal the time from, which synchronizations to make with the lines of the other performers and which to abandon.

Like some long compositions by Morton Feldman, *COALS* loses a sense of forwards and backwards motion. In the latter case at least, this is because the underlying formalisms are symmetric with respect to the direction of time.[160] But unlike Feldman's predilection for small and very large time scales, the structures in *COALS* operate primarily with moderate durations on the order of 5 to 90 seconds.

The performers of *COALS* do not have to deal with many explicitly notated indeterminacies. It occasionally happens, as with the eviolin in bars 151–152 where floor position is not specified directly but rather indicated to vary quickly through its whole range, or as with instructions to hold a note until out of

[160] The depth-first traversal of the graph in figure 43 is, strictly speaking, asymmetric with respect to time. But at this abstract level if a five-minute block of *Carta* were recomposed with paths reversed, the block could hardly be identified without prior knowledge. Even as a purely graph-theoretic exercise forward and backward paths would be almost indistinguishable.

breath. Implicitly, however, both tempo and dynamics are less determinate than one might expect. Within the four common dynamic markings of *pp*, *p*, *f*, and *ff* lies plenty of latitude for the performers to make their mark; we now discuss the origin of this dynamics system.

5.2.4 Dynamics in *Carta terrae incognitae*

Dynamics are more often point indications than crescendos or decrescendos. As the graph in figure 43 is traversed, it induces a sequence of dynamics sets (analogous to pitch sets). At each point of the graph, dynamics are chosen only from the set labeling that point's horizontal slab, one of the fifteen nonempty subsets of {*pp*, *p*, *f*, *ff*}. The boundaries between horizontal divisions in figure 43 tend towards equal numbers of points in each horizontal slab, but deliberately leave some inequality for dramatic effect. Total "democracy," a uniform distribution, would transmit the greatest amount of raw information to the listener. But more interesting are occasional reductions of information rate, so for example the unique occurrence of *tutti pp* in the second row from the bottom in figure 43 stands out in the overall dramatic flow, much more so than the half-dozen occurrences of the dynamics sets in the middle rows.

The fifteen dynamics sets are distributed among the rows to approximate uniform distributions of two readily audible properties of a dynamic set, the spread from quietest to loudest dynamic and the average dynamic. So scanning down the left column of figure 43 there is no discernible trend from, for example, louder to quieter or broad range to narrow range. In other words, position in figure 43 is strongly decorrelated with dynamic level and dynamic range. This allows for the perceptibly distinguishable construction of sections with deliberately continuous dynamic properties, by appropriately sequencing many short segments. Such continuity is unlikely to happen accidentally when the decorrelation is strong.

We have seen how the rows of figure 43 are deliberately chosen to contain different numbers of points. Thus, the dynamics sets are also distributed to strongly decorrelate (in average level and in spread) with the number of points in each row. Without this decorrelation some kinds of dynamics set (loud ones or narrow ones) would predominate at the expense of their opposites. For a composition as long as *COALS*, avoiding such an unequal distribution better maintains long-term interest. This is because a uniform distribution maximizes the rate of information flow at the global time scale on which such a distribution is perceived.

Since the sequence of dynamics sets comes from a traversal of the graph in figure 43, the onset of a new dynamics set can only occur when a new subset of instruments begins (a new point in the graph is reached). About half the time this means that a new instrument is introduced. So in rendering dynamics sets, if a new dynamic is also introduced, it is preferably assigned to the new instrument so that both structures, dynamics and timbre, are heard more clearly.

5.2.5 Pitch organization in *Carta terrae incognitae*

The several basic harmonic areas in *Carta* label the columns of figure 43. These harmonic areas we will generally call pitch sets (understood to mean pitch class sets). As with dynamics sets, the eventual order of these pitch sets in the musical surface is determined by the traversals of the graph in figure 43 and the subsequent multilayered sequencing of the paths produced.

Within each occurrence of a pitch set, the limited availability of pitch classes suggested that (as with tonal or pentatonic music) they be organized in a hierarchy. In some cases one pitch class is specified as central, and perhaps another as a “dominant” to this “tonic”. In other cases, particularly five-element pitch sets, the pitch classes are organized with a structure developed for this movement which I call a trichord series. This structure is explained in detail below.

5.2.5.1 Content of pitch sets

I wanted pitch structure to reflect timbral structure, but I could not use the distance between pitch classes directly as with the distance between timbres. This is because the distance between pitch classes is difficult to quantify pragmatically. By comparing simple frequency, two pitch classes a semitone apart are nearby while two a perfect fifth apart are almost maximally distant. But the exact opposite holds if we compare interval consonance (call it overlap of overtone structure, to safely retreat from aesthetics to objective psychoacoustics). In a large composition it is quite possible for each metric to hold at different times. Since any resolution of these opposing metrics would be complicated, I followed the conclusions of Vriend (1981) from studying elaborate techniques of Xenakis and instead chose a much simpler construction.

Twelve points were placed uniformly and randomly in a square; each point was labeled with a pitch class. Drawing the Delaunay triangulation of these points (for only twelve points, doing this by hand is straightforward) produced a graph of points connected by edges, analogous to figure 43 but having points corresponding to pitch classes instead of timbres. So the same depth-first traversal applied as before: choose a point, choose an edge leaving that point (excluding previously chosen edges) to reach another point, and keep on going until you get stuck; repeat this until you run out of edges. As it stands this traversal would produce an ordering of pitches, but instead I needed a collection of sets of pitches to apply to the columns of figure 43. So I constrained the path-construction algorithm above, similarly to how the second traversal of the graph in figure 43 was constrained: terminate the path early if five points have been visited. This produced pitch sets of size at most five; smaller ones resulted from paths which got stuck sooner.

The resulting pitch sets are listed across the top of figure 43. The assignment of pitch sets to columns derives from criteria much like that of dynamics: uniform distribution of readily audible properties such as spread and size. Pitch sets have no average value, only a spread.[161] This simplifies the assignment: pitch set average need not be decorrelated with the number of points in each column.

To maintain interest over the global duration of this movement, just as with dynamics sets the size of pitch sets is decorrelated with the number of points in each column. For example, of the two smallest pitch sets {Bb, C#} occurs only 3 times while {Eb, A} occurs 13 times. The five-element sets occur from 4 to 10 times each. So no size consistently predominates at the expense of another; this maximizes the information rate of this attribute. As with the horizontal boundaries between dynamics sets, the vertical boundaries tend towards equal numbers of points in each column but deliberately leave some inequality for dramatic effect.

5.2.5.2 Expanded pitch sets

The basic pitch sets listed in figure 43 are sometimes stretched, like the music of Brahms and his successors stretches harmony but preserve a skeleton sometimes more imagined than evident in the musical surface.

Boulez-like frequency multiplication is the simplest technique used in *Carta* to expand a pitch set. In bars 175–212 (with interruptions at bars 190–198 and 204–205) figure 43 would dictate that only the pitch set {G#, A, C} be used. Instead, this pitch set is expanded into its whole-tone flat and sharp neighbors, *i.e.*, multiplied by the pitch set {F#, A#}. This expansion is furthermore constrained so that choice of neighbor is correlated with dynamic level. Instruments playing *pp* use {F#, G#, A#}; *ff*, {A#, B, D}; *f* or *p*, the original pitch set.

A similar correlation is found in bars 108–121, where the pitch set {D, Eb, E, G, Ab} is multiplied by {C#, Eb}. The effect here is dramatic because the pitch set shifts by semitones all at once in a block

---

[161] When drawn most compactly as in figure 43, the spread of a pitch set is defined as the interval between the lowest and highest pitch. Strictly speaking, according to this definition a pitch set does have an average value, the mean between these two extrema. But this is not nearly as perceptible as the average value of a dynamics set, under normal circumstances when more than one octave is used. Pitch set average is therefore correspondingly less important to decorrelate with other parameters as a means of increasing information density.

presentation. This is a result of the concentration of dynamics sets with only one element in this passage. At the *tutti pp* in bar 111, the pitch set shifts up a semitone. Three bars later a *tutti ff* pushes it down two semitones to one below normal. Unequal dynamics return at bar 116, where *pp* and *p* are rendered one semitone sharp, *ff* and *f* one semitone flat; this correlation is reversed two bars later. Finally in bars 120–121 a *tutti ff* occurs, but now all dynamics revert to the original pitch set.

The pitch set {G#, A, C} occurs again in bars 160–172. Here *pp* and *p* instruments go up a semitone, *f* and *ff* down a semitone. A variety of dynamics combinations—all quiet, all loud, some of each—thereby produces quite a varied emphasis of intervals.

In bars 273–276 and 301–305 {G#, A, C} is multiplied by a tritone interval to make {G#, A, C, D, Eb, F#}. These passages are too short to benefit perceptibly from correlating pitch extension with other parameters.

Frequency multiplication may be too dignified a term for modifications such as the one which occurs in bars 283–285. Here eviolin and trombone briefly nudge the pitch set {G#, A, C} up a semitone, simply for variety. This brief burst of what is ostensibly {G#, A, C} occurs in the middle of a sea of {Ab, Bb, B, C, C#}, so its break with orthodoxy is not so easily noticed. (This may be what Ligeti was thinking of in applying serial techniques to error itself: an error too mild to be consciously noticed during a performance, as contrasted with more serious deviations from the formal system underlying a composition.)

Because of the narrow compositional possibilities and common occurrence of the pitch set {Eb, A} as it stands in figure 43, it is sometimes compressed in time (its predecessor and successor pitch sets invade its frontiers). This pitch set is also sometimes expanded to larger related pitch sets rather like the expansion just mentioned in bars 108–121. Figure 45 illustrates perhaps the most extreme modification. To the original set {Eb, A}, its semitone neighbors {D, E, Ab, Bb,} are added to get {D, Eb, E, Ab, A, Bb}. From this set, {Eb} is actually dropped to produce a pitch set of size five, which lets it be structured as a retrograde trichord series on {E, A, Bb, D, Ab}.

Figure 45. *Carta terrae incognitae*, bars 505–514, flute and clarinet. Expansion of the pitch set {Eb, A}. (Clarinet sounds a whole tone lower than written.)

5.2.5.3 Trichord series

Trichord series are structures in *Carta* which explicitly group different members of a pitch set and thereby generate small-scale harmonic motion as well as intervallic variety. A single trichord series is a Babbittesque sequence of all ten ways to choose three elements from a set of five, in this case all trichord subsets of a pentachord. Arranging these trichords in a sequence, operations analogous to those of dodecaphony apply. The prime form comes from the canonical ordering of choices, shown abstractly in the top row of figure 46 and in an especially transparent musical context in figure 47.

Figure 46. Trichord series on a set of five pitch classes. The four staves show the prime, retrograde, inverse, and inverse retrograde forms respectively.

The retrograde form simply reverses the order of the ten trichords (figure 46, second row). For example, bars 567–570 contain a retrograde series on {D, F#, Eb, G, F}.[162] Another retrograde series in bars 266–269 soon after the excerpt in figure 47 is on {G, F#, F, Eb, D}, the same pitch classes of the preceding series but in a different order. This strongly resembles the series in figure 47, since the retrograde series of a reversed pitch set has the same first two and last two trichords as the original. The inverse form of the series starts with the outermost two elements of the prime form and jumps back and forth until the middle is reached (figure 46, third row). Like its dodecaphonic namesake, the inverse retrograde form is both the retrograded inversion and the inverted retrogression; that is, it starts at the middle and oscillates outwards (figure 46, fourth row). Since the noninverted forms of the series have a more static harmony, more common pitch classes between successive trichords, the inverted forms are more common in *COALS*.

[162] This is a slight abuse of mathematical notation. The braces connote pitch *set*, but the set is ordered. Many different series can come from different orderings of the set. Orderings in *COALS* are determined on a case-by-case basis from the surrounding musical context. Unlike the presentation in figure 47, they often avoid consecutive elements a semitone apart in order to increase intervallic variety.

Figure 47. *Carta terrae incognitae*, bars 256–263. Unelaborated presentation of the prime form of a trichord series on {D, Eb, F, F#, G}. (Bass guitar sounds an octave lower than written.)

Trichord series can be rendered flexibly, just as with dodecaphonic series. Here are some examples. An inverted series is compressed in bars 501–503 (figure 48). Bars 285–286 contain the first nine trichords of a series; the completing tenth chord is delayed until the downbeat of bar 291, analogous to the delayed completion of a chromatic aggregate in dodecaphony. At sufficient speed, the harmonic function of a trichord series gives way to melodic functions. This is the case in bars 446–448 where cello and trombone present an unusually fast inverted series on {D, Eb, F, F#, G} (figure 49).

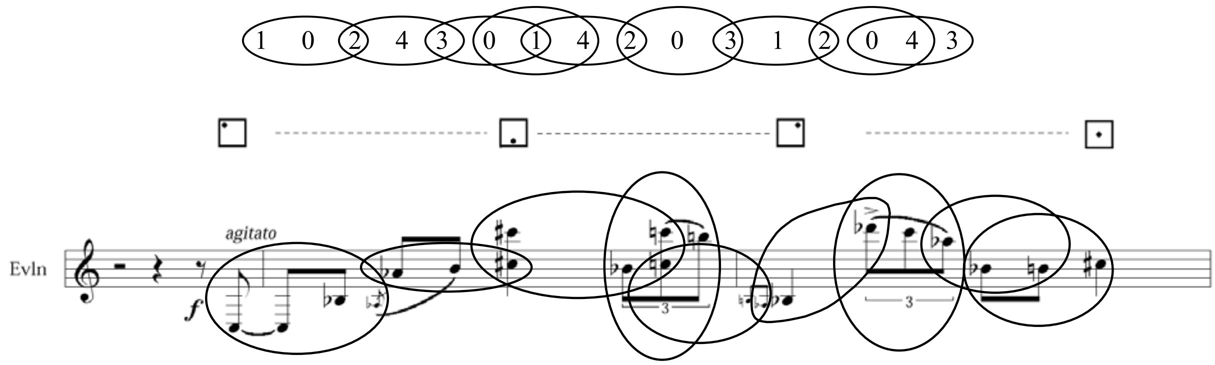


Figure 48. *Carta terrae incognitae*, bars 501–503. An inverted trichord series compressed by eliding notes common to successive trichords. Trichords are circled in the diagram. The series is built on the pitch classes {Bb, C, Ab, C#, B}.

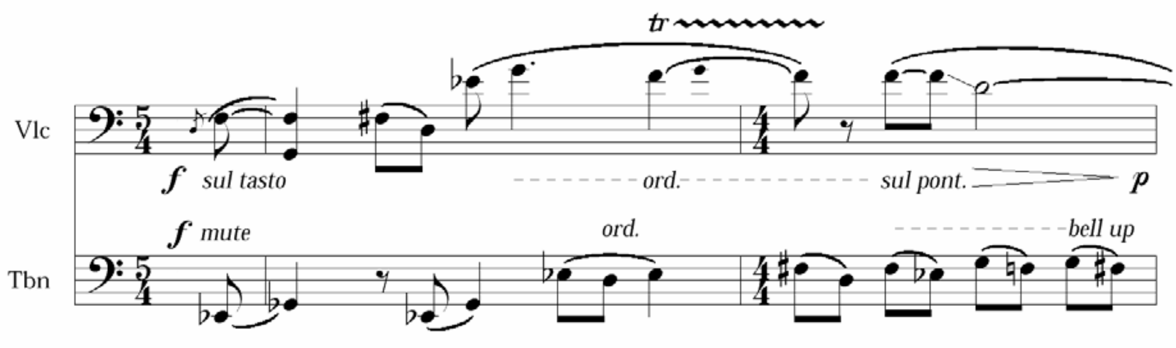


Figure 49. *Carta terrae incognitae*, bars 446–448. An unusually fast inverted trichord series on {D, Eb, F, F#, G}.

In bars 372–386 a succession of solos in flute, trombone, bass guitar and eviolin each render a trichord series in arpeggiated fashion. The same series, the prime form on {C#, B, Ab, C, Bb}, is used in all cases so these solos could even be called quasi-canonic imitation. Only one series is used because arpeggiation would obscure the series order of an inverted form beyond recognition. Pitch class counters should beware of the glissandi in the trombone part, as a glissando from C# down to B includes the pitch class C.

A concentration of trichord series is found at the very end of *Carta*. Bars 552–563 (with a short interruption) present an inverse retrograde followed by an inverted series on {Ab, Eb, G, D, E}. After a longer interruption, bars 567–573 conclude the movement with retrograde and prime series on {D, F#, Eb, G, F}.

5.2.5.4 Regions of static harmony

As with much of the music of Wolpe, the serial yet nondodecaphonic treatment of pitch in *COALS* avoids aggregate grayness. Subsets of the chromatic aggregate circulate sometimes quickly, sometimes slowly. In fact, when the longest edges of figure 43 are traversed the pitch material goes beyond coherence and approaches stasis.

The pitch set {B, C#, E, F} is the one most frequently associated with long durations. Fortuitously it contains all interval classes from semitone to tritone (an "all-interval tetrachord") so intervallic variety is maximized within this tight constraint. Even so, a few chromatic inflections like appoggiaturas are sometimes found in these passages. This is the case in the flute/eviolin/bass guitar trio in bars 122–128 and the flute and eviolin duet in bars 247–251. On the other hand, bars 190–194 strictly adhere to the pure pitch set; moreover, long chords without vibrato bring out the stasis. Emphasized stasis is desirable in this passage because it is both preceded and followed by a few dozen bars of highly active playing. (The immediately preceding bars 175–189 actually correspond to the very longest edge in figure 43, but avoid harmonic stasis by subjecting the pitch set {G#, A, C} to frequency multiplication as we have seen.) Closer to the end of the movement, bars 453–458 are again fairly pure, adding only four whole-tone (not semitone) inflections and a glissando. Stasis is brought out here by long unchanging trills and tremolos in the eviolin part.

The very first harmonically static section occurs in bars 18–31 on {Eb, A}, one of the two smallest pitch sets. A few chromatic inflections occur here, such as in bar 23 where the clarinet introduces a trumpet-like gesture which returns several times in the movement.

Finally on the pitch set {G#, A#, B, C, C#}, there are two long duets for clarinet and eviolin in bars 122–127 and near the end of the movement in bars 535–540. Since this is one of the largest pitch sets, maintaining it for an extended period reduces harmonic motion less obviously than is the case with the smaller pitch sets.

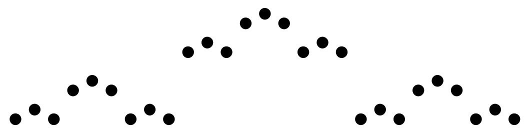
5.2.6 Motives and gestures

Short gestures are common in both movements of *COALS*: in the first by specification as we have seen, in the second by a more intuitive recognition of similar patterns being susceptible to similar motivic treatment. These little motives recur sometimes soon, sometimes much later, sometimes literally, sometimes only as an allusion not immediately recognizable. Such short gestures fit into the formal constraints more easily than long melodies. They also allow for rapid changes of texture, enabling a fluidity which is desirable when large-scale organization comes from something other than gradual Lisztian development of a few principal themes. Where Liszt orchestrates to deliver a thematic narrative, *COALS* if anything does the reverse by using motives to support and connect orchestral (timbral), dynamic, and pitch structures. As a perhaps unflattering analogy, Humphrey Bogart movies use physical action to help deliver a gripping story whereas 1990's action films like *Independence Day* and *Rumble in the Bronx* use narrative plot merely as an excuse to display heroic feats. Another medium provides a more favorable comparison: painters after Liszt's day began to work nonrepresentationally (*i.e.*, nonnarratively), leading eventually to the formal definitions of *De Stijl*, a group which influenced Boulez as we saw in chapter one. But even so, specialists can argue which was means and which was end centuries earlier, Vermeer's portrait or Vermeer's use of light and shadow.

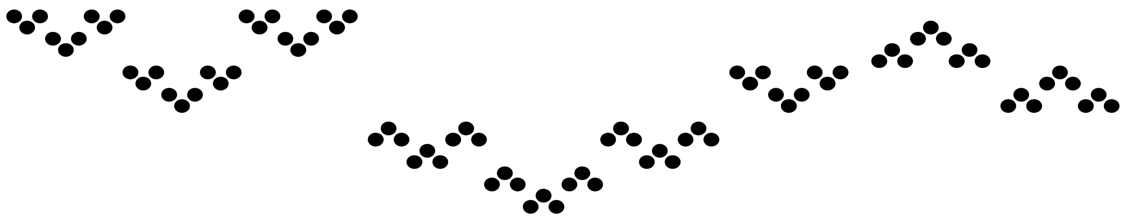
Even though motives do not play a primary structural role in *Carta*, we can consider a few examples of their recurrence in different contexts. Though separated by great distances, two easily recognizable repetitions (because of the little trumpet-call introduction) are the flute line in bars 464–467 and the clarinet in bars 146–150, which spring from the clarinet passage starting at bar 23. Even the pitch classes are preserved in bars 146–150. The immediate continuation of the clarinet line, bars 151–155, returns in the bass guitar at bars 299–302. (This continuation is also partially begun by the flute in the example just mentioned, after bar 467). Another gesture, "quasi glissando" semitone sliding, occurs early on in the clarinet, bars 44–45; it recurs in the clarinet as late as bars 536–538.

Multi-instrumental ideas also recur. The complex interplay between flute and clarinet in bars 277–280 is restated by only the flute as a compound line in bars 293–295, followed immediately in the next four bars by an amplified restatement using both flute and clarinet again. These restatements differ markedly from the original because they project it into only two pitches. The longest motivic recurrence in *Carta* is naturally enough a loose imitation and elaboration: the duet between cello and trombone in bars 129–138 is found again in the context of a different pitch set in the flute and clarinet, several minutes later in bars 273–282. In both passages the prominent duet is quietly accompanied by two other instruments.

The most expanded motivic idea occurs a little past the middle of *Carta*. In bars 321–332 the trombone pitches trace out a three-level fractal pattern, schematized in figure 50a. This is followed immediately by an inverted four-level pattern in the cello, bars 333–355 (continuing on another six bars in free imitation of this texture), figure 50b. (The bass guitar comments on the cello line, emphasizing certain notes, tracing out its own little fractal lines at times.)[163] Despite the length of these fractal processes, monotony is hardly a danger because so much else changes while the process unfolds: pitch set, dynamics, intensity, rhythmic cells, note groupings (never threes within threes), internal reflections within the fractal in the second case, and speed of presentation.



(a)



(b)

(c)

Figure 50. Fractal "melodies" in *Carta terrae incognitae.* (a) Schematic of the pitch levels in the trombone part, bars 321–332. (b) Schematic of the pitch levels in cello, bars 333–355. (c) The last third of (b) leading into free elaboration, as it appears in bars 349–355.

From late Beethoven (and many after him) *COALS* borrows the technique of long-range linear connections in pitch. For example, in bar 26 of *Camera obscura* the eviolin reaches a high F while other instruments play in less elevated tessitura. Four bars later the flute connects it to a high E, where the connection is left hanging for a while. The eviolin takes it up again briefly with a B flat in bar 44, and

[163] The parameter of register is already occupied in flute and clarinet with rendering the amount of intensity, but it is available in these lower three instruments.

only in bar 49 does the flute finally carry this line down to connect to the rest of the musical surface. Such connections occur even more frequently with low notes, since low notes have a fixed playable limit while high ones vary with performer skill. Again in *Camera obscura*, the trombone briefly visits its bottom octave and bottoms out on a G in bar 103; this register is entirely avoided until the bass guitar suddenly continues the motion downward to a single G flat in bar 122, then four bars later exactly retraces the steps of the cello back up, remembered from over a minute ago.

5.2.7 Problems of notation

Most of the notation in *COALS* is prescriptive: do this. But this is seasoned with occasional descriptive notations: make it sound like this. Examples of the latter include imprecise Italian adjectives and fermatas. (This conscious decision to combine them follows Brian Ferneyhough's example, as mentioned in chapter one.)

No extraordinary notations are used to simplify learning and rehearsal. Performers might ignore a few parameters in early rehearsals, like a choir may ignore dynamics or replace text with a neutral syllable on the first reading of a piece. The relatively ventilated texture even makes playing from full score possible since pages can be turned during rests. Playing from parts would necessitate more than the average number of cue lines, since a synchronizing pulse is for the most part absent but subtle timing relationships between instruments are common. Another practical consideration is the generous amount of vertical white space in the score. This leaves room for occasional dense accumulations of "diacritical" markings: *sul pont.*, dynamics, slurs, left-hand pizzicato markings, mute instructions, and so on.

Rhythmic notation is deliberately simple, to counter the sometimes high complexity of everything else. At least synchronization in rehearsals is made easier without challenges like overlapping, out-of-phase, or nested tuplets; multiple simultaneous tempos; or sudden jumps between precise metronome markings. Escape from the tyranny of the pulse comes through an (in some ways) easier method, grace notes. These produce deliberate local desynchronizations and occur in any of four configurations, following a note or rest and preceding a note or rest. Each grace note sounds distinctly despite its shortness: their purpose is desynchronization, not ornamentation or being secondary in any way.

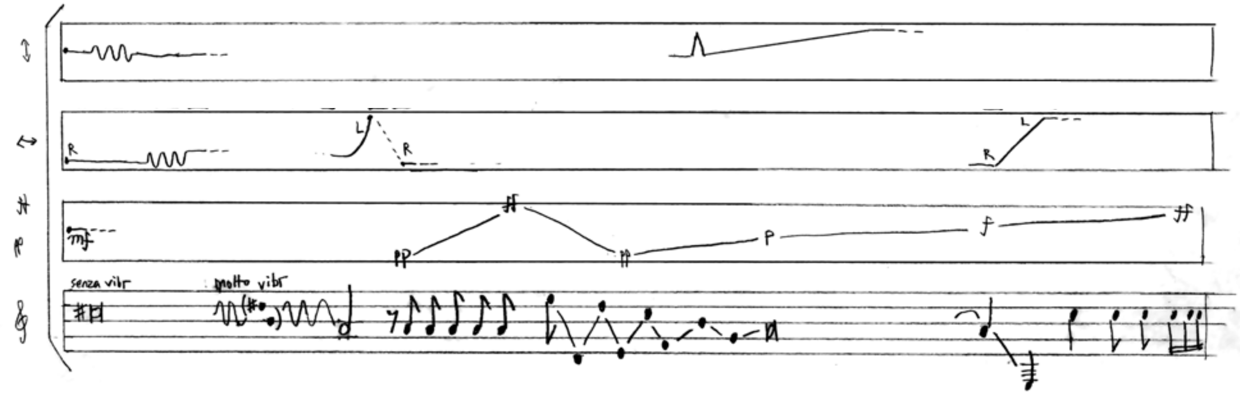
One innovation in rhythmic notation in *COALS* is the accenting of rests. An accented rest tells the performer to strongly mark the end of the preceding note. The note should not die out, but its ending should be felt as a rhythmic event in its own right. (My first organ teacher made an impression on me by pointing out that, unlike piano playing, the end of a note is just as important as the beginning.)

One extended technique used in *COALS* which has no widely accepted notation is a sequence of diminishing vertical bars on the staff, to indicate the bouncing of a bow on the strings without appreciable transverse motion.[164] This is seen for example in *Carta* bar 43, cello and bar 73, eviolin. In other cases, English phrases (footnotes) are often clearer than special notation for infrequent events. For example, for the flute overblowing at bars 9 and 11 of *Carta* a graphical depiction of the higher harmonics encountered would have been challenging to typeset and somewhat perplexing to read. The instructions to play until out of breath or out of bow in *Camera* bars 94 and 134, or how to align a grace note figure shared by flute and clarinet in *Camera* bar 116, are more efficiently communicated through prose than through clever graphic symbols. Only if one of these events occurred numerous times would a special graphical notation be more efficient for the performer to decode.

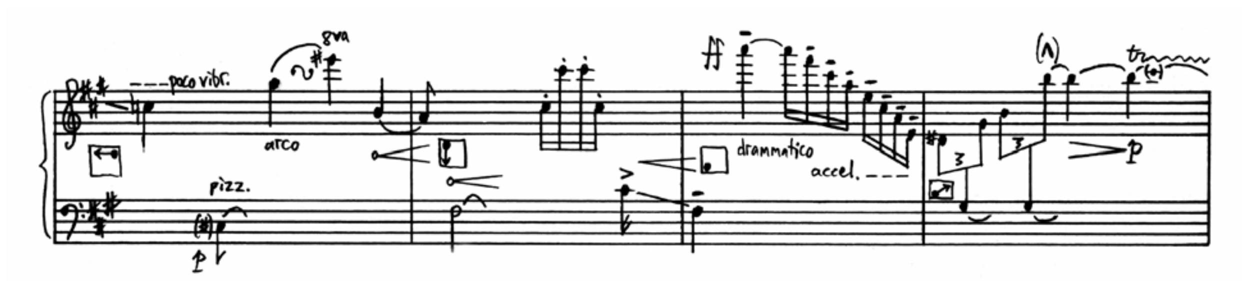
5.2.7.1 Graphical notation of spatial movement

Notating the horizontal position of the eviolin took a few tries to get it right. Strictly speaking, for eviolin this technique is not extended; but as violinists outnumber eviolinists we practically consider it to be so anyways. An early score separately graphed each parameter versus time (figure 51a). This failed to clearly communicate latitude and longitude because those are not just one but two coupled spatial parameters, and two parameters cannot fit onto one vertical axis. They need to be coupled on the score as well, to correspond to the performer's mental conception of moving around: heading for a spot on the stage, not turning two knobs on an Etch-a-Sketch. The multi-graph layout is useful for considered analysis but poor for real-time performance. Again in hindsight, a four-staff layout is difficult because human performers do not have four independently focusable eyes. Score-reading remains a multidimensional tracking task as we have seen. Fewer separate areas to look at keep it a single task instead of a multitude of unidimensional tracking tasks. In the latter case, the score cannot be read so much as memorized, one parameter at a time if necessary.

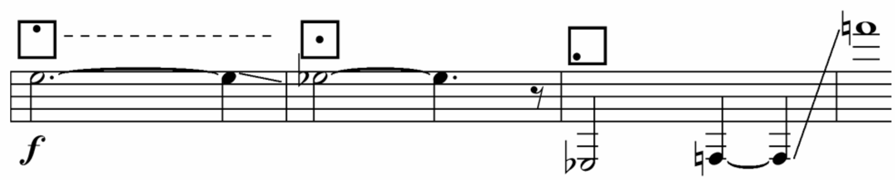
[164] This notation, apparently originated by Benjamin Britten, was brought to my attention by David Bohn.



(a)



(b)



(c)

Figure 51. (a) Separate graphs of each parameter versus time. (b) Dot-in-square notation, including direction arrows. (c) Dot-in-square notation as used in *COALS* (*Camera obscura*, bars 131–134, eviolin).

So with reduced spatial resolution but increased clarity and concision, a later study patterned indications of spatial position after indications of dynamics (figure 51b): small glyphs inserted at points in the score. These square glyphs were either *subito* or contained arrows analogous to *p crescendo* to indicate future direction of movement. These arrows turned out to be more cluttering than helpful so they were replaced in *COALS* with dotted lines connecting squares, analogous to a hairpin marking between *p* and *f*. This final notation is shown in figure 51c.

List of References

Agronin, M. 1987. The design of a nine-string six-degree-of-freedom force-feedback joystick for telemanipulation. *Proc. NASA Workshop on Space Telerobotics* 1987:341–348.

Aikin, J. 1990a. Buchla Thunder MIDI performance interface. *Keyboard* 16(7):124–135. San Francisco: Miller Freeman Publications.

———. 1990b. DrumKat MIDI percussion controller. *Keyboard* 16(7):136-142. San Francisco: Miller Freeman Publications.

Airey, J., J. Rohl, and F. Brooks Jr. 1990. Towards image realism with interactive update rates in complex virtual building environments. *Interactive 3D* 1990: 41–50.

Allen, J. 2000. *Notebender Keyboard with longitudinal key motion—photo and description.* <http://www.bikexprt.com/music/notebend1.htm>.

Allison, L. 1984. *Dynamic Programming Algorithm, Edit Distance.* <http://www.csse.monash.edu.au/%7Elloyd/tildeAlgDS/Dynamic/Edit.html>.

Alternate Mode, Inc. 2000. *KAT Electronic percussion*. <http://www.katpercussion.com/>.

Babbitt. M. 1955. Some aspects of twelve-tone composition. *The Score* XII:53–61.

———. 1958. Who cares if you listen? *High Fidelity* VIII(2):38–40, 126–127. Reprinted in *Contemporary Composers on Contemporary Music*, expanded edition, ed. E. Schwarz and B. Childs, 244–250. New York: Da Capo, 1998.

———. 1962. Twelve-tone rhythmic structure and the electronic medium. *Perspectives of New Music* 1(1):49–79.

———. 1972a. Past and present concepts of the nature and limits of music. In *Perspectives on Contemporary Music Theory*, ed. B. Boretz and E. Cone. New York: W.W. Norton.

———. 1972b. The structure and function of musical theory. In *Perspectives on Contemporary Music Theory*, ed. B. Boretz and E. Cone. New York: W.W. Norton.

Baker, M. P., and C. Wickens. 1995. *Human factors in virtual environments for the visual analysis of scientific data*. Technical report, National Center for Supercomputing Applications, Urbana, Illinois.

Bargar, R., I. Choi, S. Das, and C. Goudeseune. 1994. Model-based interactive sound for an immersive virtual environment. *Proc. Intl. Computer Music Conf.* 1994:471–474.

Barrière, J.-B., Y. Potard, and P.-F. Baisnée. 1985. Models of continuity between synthesis and processing for the elaboration and control of timbre structures. *Proc. Intl. Computer Music Conf.* 1985:193–198.

Bauer, W., and B. Foss. 1992. GAMS: an integrated media control system. *Computer Music J.* 16(1):19–24.

Behrman, D. 1976. What indeterminate notation determines. In *Perspectives on Notation and Performance*, ed. B. Boretz and E. Cone, 74–89. New York: W.W. Norton.

Bell, J., M. Berger, J. Saltzman, and M. Welcome. 1994. A three dimensional adaptive mesh refinement for hyperbolic conservation laws. *SIAM J. Sci. Computing* 15:127–138.

Berger, M., and P. Colella. 1989. Local adaptive mesh refinement for shock hydrodynamics. *J. Comp. Phys.* 82:64–84.

Berger, M., and J. Oliger. 1984. Adaptive mesh refinement for hyperbolic partial differential equations. *J. Comp. Phys.* 53: 484-512.

Bezdek, J. 1993. Fuzzy models—what are they, and why? *IEEE Trans. Fuzzy Systems* 1(1):1–6.

Blevins, R. 1987. Yamaha digital synthesis and controller products. *Computer Music J.* 11(4):59.

Boie, R., M. Mathews, and A. Schloss. 1989. The radio drum as a synthesizer controller. *Proc. Intl. Computer Music Conf.* 1989:42–45.

Bosanquet, R. 1875. Temperament; or, the division of the octave. *Proc. Musical Assn.* 1:4–17,112-158. Reprint. Vaduz: Kraus, 1966.

Boulez, P. 1954. Recherches maintenant. Reprinted in *Relevés d'apprenti*, ed. P. Thevenin, 27–32. Paris: Seuil, 1966.

———. 1955. An der Grenze des Fruchtlandes. *Die Reihe* 1:47–57. Universal Edition, Vienna.

———. 1963a. *Boulez on Music Today.* Trans. S. Bradshaw and R. Bennett. Cambridge: Harvard Univ. Press, 1971.

———. 1963b. Dire, jouer, chanter. *La musique et ses problèmes contemporains 1953–1963*. Paris: Juillard.

———. 1964. Alea. Trans. D. Noakes and P. Jacobs. *Perspectives of New Music* 3(1):42–53.

———. 1976. *Conversations with Célestin Deliège.* Eulenberg Books, London.

Bowler, I., P. Manning, A. Purvis, and N. Bailey. 1990. On mapping *n* articulation onto *m* synthesizer-control parameters. *Proc. Intl. Computer Music Conf.* 1990:181–184.

Brady, R., J. Pixton, G. Baxter, P. Moran, C. Potter, B. Carragher, and A. Belmont. 1995. "Crumbs: a virtual environment tracking tool for biological imaging." *Proc. IEEE Symposium on Frontiers in Biomedical Visualization* 1995:18–25.

Bregman, A. 1990. *Auditory Scene Analysis*. Cambridge, Mass.: M.I.T. Press.

Bromwich, M. 1997. The Metabone: An interactive sensory control mechanism for virtuoso trombone. *Proc. Intl. Computer Music Conf.* 1997:473–475.

Brown, E., W. Buxton, and K. Murtagh. 1990. Windows on tablets as a means of achieving virtual input devices. *Human-Computer Interaction – INTERACT* 1990:675–681. Amsterdam: Elsevier Science Publishers, ed. D. Diaper et al.

Bryan, B. 1999. Surfaces and TINs. Lecture notes for the course *Geographic Information Systems III*. <http://www.gisca.adelaide.edu.au/~bbryan/lectures/surfaces_tins/>.

Buxton, W. 1983. Lexical and pragmatic issues of input structures. *Computer Graphics* 17(1): 31–37.

———. 1986. There's more to interaction than meets the eye: some issues in manual input. In *User Centered System Design: New Perspectives on Human-Computer Interaction*, ed. Norman and Draper, 319–337. Hillsdale, New Jersey: Erlbaum.

Buxton, W., R. Hill, and P. Rowley. 1985. Issues and techniques in touch-sensitive tablet input. *Proc. SIGGRAPH* 1985:215–224.

Cadoz, C., L. Lisowski, and J-L. Florens. 1990. Modular feedback keyboard. *Proc. Intl. Computer Music Conf.* 1990:379–382.

Cai, Y., S. Wang, M. Ishii, R. Takamatsu, and M. Sako. 1995. 3-D force display device for big virtual space. *VSMM Intl.* 1995:171–176.

Cariou, B. 1994. The aXi0 MIDI Controller. *Proc. Intl. Computer Music Conf.* 1994:163–166.

Chafe, C. 2000. Personal communication.

Chafe, C., and S. O'Modhrain. 1996. Musical muscle memory and the haptic display of performance nuance. *Proc. Intl. Computer Music Conf.* 1996:428–431.

Chase, W., and H. Simon. 1973. Perception in chess. *Cognitive Psychology* 4:55–81.

Chen, E. 1999. Six degree-of-freedom haptic system for desktop virtual prototyping applications. *Proc. First Intl. Workshop on Virtual Reality and Prototyping*: 97–106, Laval, France.

Choi, I. 1997. Interactivity vs. control: human-machine performance basis of emotion. *Kansei—the Technology of Emotion: AIMI Intl. Workshop Proc.* 1997:24–35.

Choi, I., R. Bargar, and C. Goudeseune. 1995. A manifold interface for a high dimensional control space. *Proc. Intl. Computer Music Conf.* 1995:385–392.

Clarke, M. 1999. Composing with multi-channel spatialisation as an aspect of synthesis. *Proc. Intl. Computer Music Conf.* 1999:17–19.

Clarkson, K. 1992. Randomized geometric algorithms. In *Computing in Euclidean Geometry*, 117–162. River Edge, New Jersey: World Sci. Publishing.

Clarkson, K., K. Mehlhorn, and R. Deidel. 1993. Four results on randomized incremental constructions. In *Computational Geometry: Theory and Applications*, 185–221. <http://cm.bell-labs.com/netlib/voronoi/hull.html>.

Collins, F., Jr., and P. Bolstad. 1996. A comparison of spatial interpolation techniques in temperature estimation. *Proc. 3rd Intl. Conf./Workshop on Integrating GIS and Environmental Modeling*, Santa Fe. Santa Barbara: National Center for Geographic Information and Analysis. <http://www.ncgia.ucsb.edu/conf/SANTA_FE_CD-ROM/sf_papers/collins_fred/collins.html>.

Cook, P. 1991. *Identification and control of parameters in an articulatory vocal tract model, with applications to the synthesis of singing*. Ph.D. dissertation, Dept. of Elec. Eng., Stanford Univ.

———. 1992. A meta-wind-instrument physical model, and a meta-controller for real time performance control. *Proc. Intl. Computer Music Conf.* 1992:273–276.

———. 1993. A MIDI control and performance system for brass instruments. *Proc. Intl. Computer Music Conf.* 1993:130–133.

Cook, P., and G. Scavone. 2000. *The Synthesis Toolkit (STK)*. <http://ccrma-www.stanford.edu/CCRMA/Software/STK/>.

Cope, D. 1971. *New directions in music*. Dubuque, Iowa: Wm. C. Brown.

Cruz-Neira, C., D. Sandin, and T. DeFanti. 1993. Virtual reality: the design and implementation of the CAVE. *Proc. SIGGRAPH* 1993:135–142.

Cuzzucoli, G., and V. Lombardo. 1997. Physical model of the plucking process in the classical guitar. *Proc. Intl. Computer Music Conf.* 1997:172–179.

Dawe, G. 2000. Personal communication.

Dean, T., and Wellman, M. 1991. *Planning and control*. San Mateo, California: Morgan Kaufmann.

DeLio, T. 1980. Iannis Xenakis' *Nómos Alpha*: the dialectics of structure and materials. *J. Music Theory* 24(1):63–95.

Denes, J. 1974. *Latin squares and their applications.* New York: Academic Press.

Dodge, C., and T. Jerse. 1984. *Computer music*. New York: Schirmer.

Duringer, J. 1996. The Monolith: a two-dimensional keyboard for pitch and timbre. *Experimental Musical Instruments* 12(1).

Dwyer, R. 1991. Higher-dimensional Voronoi diagrams in linear expected time. *Discrete and Computational Geometry* 6(4):343–367.

Eckstein, B. 1989. Evaluation of spline and weighted average interpolation algorithms. *Computers and Geosciences* 15(1):79–94.

Edahiro, M., I. Kokubo, and T. Asano. 1984. A new point-location algorithm and its practical efficiency—comparison with existing algorithms. *ACM Trans. on Graphics* 3(2):86–109.

Favilla, S. 1994. The LDR controller. *Proc. Intl. Computer Music Conf.* 1994:177–180.

Feiten, B., and S. Günzel. 1993. Distance measure for the organization of sounds. *Acustica* 78(3):181–184.

Feldman, J. 1983. *Learning tabla with Alla Rakha: a manual of drumming and recitation*. Los Angeles: Ravi Shankar Music Circle.

Feller, R. 1994. *Multicursal labyrinths in the work of Brian Ferneyhough*. D.M.A. Thesis, Univ. of Illinois, Urbana-Champaign.

Fels, S., and G. Hinton. 1998. Glove-Talk II: a neural network interface which maps gestures to parallel formant speech synthesizer controls. *IEEE Trans. Neural Networks* 9(1):205–212.

Fitts, P. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *J. Exper. Psych.* 47:381–391.

Fitz, K. 2001. *Loris*. <http://www.sourceforge.net/projects/loris>.

Fitz, K., L. Haken, and P. Christensen. 2000. A new algorithm for bandwidth association in bandwidth-enhanced additive sound modeling. *Proc. Intl. Computer Music Conf.* 2000:384–387.

Franco, S. 1974. *Hardware design of a real-time musical system*. Ph.D. dissertation, Univ. of Illinois at Urbana-Champaign.

Freed, A. 2000a. *Musical applications of resonance models.* <http://cnmat.cnmat.berkeley.edu/CAST/Server/timbralprotos.html#resonance_proto>.

———. 2000b. *Resonance models.* <http://cnmat.cnmat.berkeley.edu/Research/Resonances>.

Garner, W. 1962. *Uncertainty and structure as psychological concepts*. New York: Wiley.

Garnett, G., and C. Goudeseune. 1999. Performance factors in control of high-dimensional spaces. *Proc. Intl. Computer Music Conf.* 1999:268–271.

Gest, T., and J. Schlesinger. 1995. *MedCharts anatomy*. New York: ILOC.

Gillespie, B. 1992. The Touchback Keyboard. *Proc. Intl. Computer Music Conf.* 1992:447–448.

Gillett, R., K. Smith, and B. Pritchard. 1985. MADDM—Dance-directed music. *Proc. Intl. Computer Music Conf.* 1985:329-330.

Gold, C. 1989. Surface interpolation, spatial adjacency and geographical information systems. In *Three Dimensional Applications in Geographic Information Systems*, ed. J. Raper, 21–35. London: Taylor and Francis.

———. 1990. Space revisited—back to the basics. *Proc. Intl. Symposium on Spatial Data Handling* 4:175–189, Zurich.

Goudeseune, C., and H. Kaczmarski. 2001. Composing outdoor augmented-reality sound environments. *Proc. Intl. Computer Music Conf.* 2001.

Grey, J. 1975. *An exploration of musical timbre*. Center for Computer Research in Music and Acoustics, Dept. of Music Report No. STAN-M-2, Stanford Univ.

Griffiths, P. 1978. *Boulez*. London: Oxford Univ. Press.

———. 1994. *Modern music: a concise history*. Revised edition. Malta: Interprint.

Haken, L, K. Fitz, E. Tellman, P. Wolfe, and P. Christensen. 1997. A continuous music keyboard controlling polyphonic morphing using bandwidth-enhanced oscillators. *Proc. Intl. Computer Music Conf.* 1997:375–378.

Haken, L., E. Tellman, and P. Wolfe. 1998. An indiscrete music keyboard. *Computer Music J.* 22(1):30–48.

Hamman, M. 1997a. Composition of data and process models: a paralogical approach to human/computer interaction. *Proc. Intl. Computer Music Conf.* 1997:109–112.

———. 1997b. *Computer music composition and the dialectics of interaction*. D.M.A. Thesis, Univ. of Illinois, Urbana-Champaign.

Hardy, R. 1990. Theory and applications of the multiquadric-biharmonic method. *Computers and Mathematics with Applications* 19:163–208.

Harenberg, M. 2000. *Die Rationalisierung des Virtuellen in der Musik*. <http://audio.uni-lueneburg.de/Text/ratio.html>.

Hollis, J. 1998. J. Hollis. *SynthAxe: advanced MIDI guitar controller*. <http://www.hollis.co.uk/john/synthaxe.html>.

von Hornbostel, E., and C. Sachs. 1914. Systematik der Musikinstrumente: ein Versuch. *Zeitschrift für Ethnologie* 46(4-5):553–590, Berlin.

———. 1961. Classification of musical instruments. Trans. A. Baines and K. Wachsmann. *Galpin Soc. J.* 14:3–29.

Hunt, A., and R. Kirk. 1999. Radical user interfaces for real-time control. *Proc. Euromicro 1999*, 2:6–12. Los Alamitos, Calif.: IEEE Computer Soc.

Hutchinson, M., and P. Gessler. 1994. Splines—more than just a smooth interpolator. *Geoderma* 62:45–67.

InterSense Inc. 2001. *InterSense Inc*. <http://www.isense.com>.

Jehan, T. 2000. Personal communication.

Jensen, E. 1998. *Jensen electric violin user's manual*. Jensen Musical Instruments, Seattle.

Johnstone, E. 1985. The Rolky: a poly-touch controller for electronic music. *Proc. Intl. Computer Music Conf.* 1985:291–295.

Kajitani, M. 1992. Simulation of musical performances. *J. of Robotics and Mechatronics* 4(6):462–465.

Kaper, H., and S. Tipei. 1998. Manifold compositions, music visualization, and scientific sonification in an immersive virtual-reality environment. *Proc. Intl. Computer Music Conf.* 1998:399–405.

———. 1999. Formalizing the concept of sound. *Proc. Intl. Computer Music Conf.* 1999:387–390.

Keislar, D. 1987. History and principles of microtonal keyboards. *Computer Music J.* 11(1):18–28.

Kelley, C. 1968. *Manual and automatic control: a theory of manual control and its application to manual and to automatic systems*. New York: Wiley.

Koblyakov, L. 1990. *Pierre Boulez: a world of harmony*. New York: Academic Publishers.

Kohonen, T. 1997. *Self-organizing maps*, 2nd ed. Berlin: Springer.

Krige, D. 1981. Lognormal-de Wijsian geostatistics for ore evaluation. *South African Institute of Mining and Metallurgy Monograph Series: Geostatistics I.* South Africa Institute of Mining and Metallurgy, Johannesburg.

Kuhn, W. 1990. A real-time pitch recognition algorithm for music applications. *Computer Music J.* 14(3):60–71.

Kvifte, T. 1989. *Instruments and the electronic age: toward a terminology for a unified description of playing technique*. Oslo: Solum Forlag.

Langmead, C. 1995. *A theoretical model of timbre perception based on morphological representations of time-varying spectra*. M.A. Thesis, Dartmouth College, Hanover, New Hampshire.

de Laubier, S. 1998. The Meta-Instrument. *Computer Music J.* 22(1):25–29.

Lawson, C. 1977. "Software for C^1 surface interpolation." *Mathematical Software III*, ed. J. Rice. Academic Press, New York.

Lerdahl, F., and R. Jackendoff. 1983. *A generative theory of tonal music*. Cambridge, Mass.: M.I.T. Press.

Levenshtein, V. 1966. Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics Doklady* 10(8):707-710.

Ligeti, G. 1958. Pierre Boulez. *Die Reihe* 4:36–62. Trans. L. Black. Bryn Mawr, Penn.: Theodore Presser, 1960.

Livingston, M. 2000. Personal communication.

Livingston, M., and A. State. 1997. Magnetic tracker calibration for improved augmented reality registration. *Presence: Teleoperators and Virtual Environments* 6(5):532–546. Cambridge, Mass.: M.I.T. Press.

Lorrain, D. 1980. A panoply of stochastic 'cannons'. *Computer Music J.* 4(1):53–81.

Lund, E. 1996. Lecture on models in composition. Music 422, Univ. of Illinois, Urbana.

Machover, T. 1997. *"Classic" Hyperinstruments 1986–1992: a composer's approach to the evolution of intelligent musical instruments*. <http://kronos.media.mit.edu/Archive/Hyperinstruments/classichyper.html>.

Maconie, R. 1976. *The works of Karlheinz Stockhausen*. New York: Oxford Univ. Press.

Maeda, S., and K. Honda. 1993. From EMG to formant patterns of vowels: the implication of vowel systems and spaces. *ATR Technical Report TR-H-020*. Kyoto: ATR Human Information Processing Research Laboratories.

Mahillon, V.-C. 1893. *Catalogue descriptif et analytique du Musée instrumental du Conservatoire royal de musique de Bruxelles*. Brussels : Th. Lombaerts.

Maldonado, G. 1999. VMCI, an interface to control Csound in real-time under Windows. *Proc. Intl. Computer Music Conf.* 1999:307–310.

Martinetz, T., and K. Schulten. 1993. *Competitive Hebbian rule forms Delaunay triangulations, perfect topology preserving maps, and discrete models of manifolds*. Theoretical Biophysics Technical Report UIUC-BI-TB-93-04. Urbana: The Beckman Institute, Univ. of Illinois.

Martirano, J. 1998. *SAL-MAR Construction*. <http://ems.music.uiuc.edu/~martiran/HTdocs/salmar.html>.

Mathews, M. 1969. *The technology of computer music*. Cambridge, Mass.: M.I.T. Press.

Mead, A. 1994. *An introduction to the music of Milton Babbitt*. Princeton, New Jersey: Princeton Univ. Press.

Messiaen, O. 1942. Petite théorie de mon langage rythmique. Introduction to *Quatuor pour la Fin du Temps*. Paris: Durand.

MIDI Manufacturers Association. 1996. *The complete MIDI 1.0 detailed specification.* Los Angeles.

Miner, N., B. Gillespie, and T. Caudell. 1996. Examining the influence of audio and visual stimuli on a haptic display, *IMAGE 1996 Conf. Proc.*, Phoenix, Arizona.

Mitas, L., and H. Mitasova. 1999. Spatial interpolation. In *Geographical Information Systems: Principles, Techniques, Management and Applications*, ed. P. Longley, M. Goodchild, D. Maguire, and D. Rhind, 481–492. Cambridge: GeoInformation Intl.

Mitasova, H., L. Mitas, W. Brown, D. Gerdes, I. Kosinovsky, and T. Baker. 1995. Modelling spatially and temporally distributed phenomena: new methods and tools for GRASS GIS. *Intl. J. Geog. Info. Sys.* 9:433–446.

Mitra, S., M. Parashar, and J. Browne. 1998. Supercomputing '97: tutorial slides. *DAGH: User's Guide*. <http://www.cs.utexas.edu/users/dagh/welcome.html>.

Moog, R. 1982. A multiply touch-sensitive clavier for computer music. *Proc. Intl. Computer Music Conf.* 1982:155–159.

———. 1987. Position and force sensors and their application to keyboards and related control devices. *Proc. Audio Eng. Soc.* 1987:173–181.

Morton, A. 2001. Linux scheduling latency. <http://www.uow.edu.au/~andrewm/linux/schedlat.html>.

Nejad, O. 1998. Virtual Director. <http://www.sfgate.com/hypertek/9812/2-virtualdirector.shtml>.

Northcott, B. 1971. Boulez's theory of composition. *Music and Musicians* 20(4):32–36.

Oliver, M., and R. Webster. 1990. Kriging: a method of interpolation for geographical information systems. *Intl. J. Geog. Info. Sys.* 4(3):313–332.

O'Modhrain, M., and B. Gillespie. 1997. The Moose: A haptic user interface for blind persons. *Proc. 6^{th} World Wide Web Conf.: Access Track*.

Paradiso, J. 1997. Electronic music interfaces. *IEEE Spectrum* 34(12).

———. 1998. Electronic music interfaces. <http://www.spectrum.ieee.org/select/1297/muse.html#s1>.

———. 1999a. The Digital Baton. <http://www.media.mit.edu/~joep/SpectrumWeb/captions/Baton.html>.

———. 1999b. *The Magic Carpet*. <http://www.media.mit.edu/~joep/SpectrumWeb/captions/Carpet.html>.

———. 1999c. *Yo-Yo Ma and the Hypercello*. <http://www.media.mit.edu/~joep/SpectrumWeb/captions/Cello.html>.

Pellegrino, R. 2000. Music visualizers. <http://www.microweb.com/ronpell/MusicVisualizers.html>.

Perle, G. 1977. *Serial composition and atonality: an introduction to the music of Schoenberg, Berg, and Webern*. 4^{th} edition, revised. Los Angeles: Univ. of California Press.

Potard, Y., P.-F. Baisnée, and J.-B. Barrière. 1986. Experimenting with models of resonance produced by a new technique for the analysis of impulsive sounds. *Proc. Intl. Computer Music Conf.* 1986:269–274.

Puckette, M., T. Apel, and D. Zicarelli. 1998. Real-time audio analysis tools for Pd and MSP. *Proc. Intl. Computer Music Conf.* 1998:109–112.

Rhea, T. 1989. Clara Rockmore: the art of the theremin. *Computer Music J.* 13(1):61–63.

Ricci, C. 2000. Personal communication.

Roads, C. 1987. Big Briar multiply-touch-sensitive keyboards. *Computer Music J.* 11(4):59.

Rodet, X. 1984. Time domain formant-wave-function synthesis. *Computer Music J.* 8(3):9-14.

Rokeby, D. 1995. Transforming mirrors: subjectivity and control in interactive media. *Critical Issues in Electronic Media*, ed. S. Penny. SUNY Press, Albany, New York.

Rovan, J., and V. Hayward. 2000. Typology of tactile sounds and their synthesis in gesture-driven computer music performance. In *Trends in gestural control of music*, ed. M. Wanderley and M. Battier, 355–368. Paris: Ircam Centre Pompidou.

Rosenboom, D. 1976. *Biofeedback and the arts: results of early experiments*. Vancouver: A.R.C. Publications.

———. 1990. *Extended musical interface with the human nervous system: assessment and prospectus*. Berkeley: Intl. Soc. for the Arts, Sciences and Technology.

Ryan, J. 1992. Effort and expression. *Proc. Intl. Computer Music Conf.* 1992:414–416.

Rynes, E. 1999. Personal communication.

Sambridge, M., J. Braun, and H. McQueen. 1995. Geophysical parametrization and interpolation of irregular data using natural neighbors. *Geophysical J. Intl.* 122:837–857.

Sárközy, F. 1998. *GIS functions—interpolation*. <http://www.agt.bme.hu/public_e/funcint/funcint.html>.

Schaeffer, P. 1966. *Traité des objets musicaux*. Paris: Seuil.

Schelleng, J. 1973. The bowed string and the player. *J. Acoust. Soc. Amer.* 53(1):26–41.

Schloss, W. 1990. Recent advances in the coupling of the language Max with the Mathews/Boie Radio Drum. *Proc. Intl. Computer Music Conf.* 1990:398–400.

Secor, G. 1975. Specifications of the Motorola Scalatron. *Xenharmonicon* 2(2).

SensAble Technologies, Inc. 2000. *SensAble Technologies.* <http://www.sensable.com>.

Serafin, S., R. Dudas, M. Wanderley, and X. Rodet. 1999. Gestural control of a real-time physical model of a bowed string instrument. *Proc. Intl. Computer Music Conf.* 1999:375–378.

Sharma, R., I. Poddar, E. Ozyildiz, S. Kettebekov, H. Kim, and T. S. Huang. Toward interpretation of natural speech/gesture for spatial planning on a virtual map. *Proc. Advanced Display Federated Lab. Symp.* 1999:35–39.

Shepard, D. 1968. A two-dimensional interpolation function for irregularly spaced data. *Proc. Natl. Conf. ACM* 23:517–524.

Sheridan, T., and W. Ferrell. 1981. *Man-machine systems: information control and decision models of human performance*. Cambridge, Mass.: M.I.T. Press.

Shibuya, K., Y. Chikaoka, T. Koyama, and S. Sugano. 1997. The planning of violin playing robot with KANSEI information—algorithm to decide bowing parameters from timbre information. *6th IEEE Intl. Workshop on Robot and Human Communication*. Sendai, Japan.

Sibson, R. 1980. The Dirichlet tessellation as an aid in data analysis. *Scand. J. Stat.* 7:14–20.

———. 1982. A brief description of natural neighbour interpolation. In *Interpreting Multivariate Data*, ed. V. Barnett, 21–26. London: Wiley.

Siegel, W. 1998. DIEM: studio report. *Proc. Intl. Computer Music Conf.* 1998:312–314.

Siegel, W., and J. Jacobsen. Composing for the Digital Dance Interface. *Proc. 1999 Intl. Computer Music Conf.* 1999:276–277.

Simpson, J., and E. Weiner, eds. 1989. *The Oxford English Dictionary*. 2nd edition. Oxford: Oxford Univ. Press.

Smith, S., D. Duke, T. Marsh, M. Harrison, and P. Wright. 1998. Modelling interaction in virtual environments. *Proc. UK-VRSIG*.

Snell, J. 1983. Sensors for playing computer music with expression. *Proc. Intl. Computer Music Conf.* 1983:113–126.

Spaceball Technologies Inc. 1993. *SpaceWare programmer's reference manual*. Lowell, Mass.

Stacey, P. 1987. *Boulez and the modern concept.* Lincoln: Univ. of Nebraska Press.

Starr Labs. 1998a. *MicroZone microtonal keyboard*. <http://www.starrlabs.com/uzone.htm>.

———. 1998b. *Starr Labs*. <http://www.catalog.com/starrlab/>.

Stockhausen, K. 1955. Aktuelles. *Die Reihe* 1:57–63. Vienna: Universal Edition.

———. 1957. ...wie die Zeit vergeht... *Die Reihe* 3:13–42. Vienna: Universal Edition.

———. 1962. The concept of unity in electronic music. *Perspectives of New Music* 1(1):39–48.

———. 1989. *Stockhausen on music: lectures and interviews*. London: Marion Boyars.

Tactex Controls. 1999. *Tactex Controls.* <http://www.tactex.com>.

Tarabella, L. 1997. Studio report of the computer music lab of CNUCE/C.N.R. *Proc. Intl. Computer Music Conf.* 1997:86–88.

Théberge, P. 1997. *Any sound you can imagine: making music / consuming technology.* London: Wesleyan Univ. Press.

Thiébaux, M. 1997. *Virtual Director: steering scientific visualization with virtual camera choreography*. Master's Thesis, Dept. of Elec. Eng. and Comp. Sci., Univ. of Illinois at Chicago.

Tipei, S. 1981. Solving specific compositional problems with MP1. *Proc. Intl. Computer Music Conf.* 1981:101–109.

———. 1999. Personal communication.

Trueman, D. 1998. *The Trueman-Cook R-Bow*. <http://www.music.princeton.edu/~dan/rbow/rbow.html>.

———. 2000. Personal communication.

Trueman, D., and P. Cook. 1999. BoSSA: the deconstructed violin reconstructed. *Proc. Intl. Computer Music Conf.* 1999:232–239.

Ungvary, T., and R. Vertegaal. 1999. The SensOrg: time-complexity and the design of a musical cyberinstrument. *Proc. Intl. Computer Music Conf.* 1999:363–366.

Vercoe, B. 1979. *Reference manual for the MUSIC 11 sound synthesis language*. Cambridge, Mass.: Experimental Music Studio, M.I.T.

Vertegaal, R., B. Eaglestone, and M. Clarke. 1994. An evaluation of input devices for use in the ISEE human-synthesizer interface. *Proc. Intl. Computer Music Conf.* 1994:159–162.

Vertegaal, R., T. Ungvary, and M. Kieslinger. 1996. Towards a musician's cockpit: transducers, feedback and musical function. *Proc. Intl. Computer Music Conf.* 1996:308–311.

Vriend, J. 1981. 'Nomos Alpha' for violoncello solo (Xenakis 1966). Analysis and comments. *Interface* 10(1):15–82.

Waisvisz, M. 1985. The HANDS, a set of remote MIDI controllers. *Proc. Intl. Computer Music Conf.* 1985:313–318.

Walker, W. 1994. *A conversation-based framework for musical improvisation*. Ph.D. dissertation, Univ. of Illinois at Urbana-Champaign.

Watson, D. 1992. *Contouring: a guide to the analysis and display of spatial data*. New York: Pergamon Press.

Westergaard, P. 1963. Webern and 'total organization': an analysis of the second movement of the Piano Variations, Op. 27. *Perspectives of New Music* 1(2):107–120.

———. 1965. Some problems raised by the rhythmic procedures in Milton Babbitt's *Composition for Twelve Instruments*. *Perspectives of New Music* 4(1):109–118.

Wilson, E. 1983. *Some basic patterns underlying genus 12 and 17*. <http://www.anaphoria.com/genus.PDF>.

Wilson-Bokowiec, J., and M. Bromwich. 1999. *The Bodycoder system: intensifying the collaborative process through the use of new technology*. Osnabrück, Germany: KlangArt Congress. <http://www.hud.ac.uk/schools/music+humanities/music/KlangPaper.html>.

Winkler, T. 1995. Making motion musical: gesture mapping strategies for interactive computer music. *Proc. Intl. Computer Music Conf.* 1995:261–264.

———. 1997. Creating interactive dance with the Very Nervous System. *Proc. Connecticut College Symposium on Art and Technology*.

Wolpe, S. 1959. Thinking Twice. In *Contemporary composers on contemporary music*, expanded edition, ed. E. Schwarz and B. Childs. New York: Da Capo, 1998.

Wright, M. 2000. *OpenSound Control*. <http://cnmat.cnmat.berkeley.edu/OSC/>.

Wuorinen, C. 1979. *Simple Composition.* New York: Longman.

Xenakis, I. 1955. La crise de la musique serielle. *Gravesaner Blätter* 1(2):2–4.

———. 1990. Sieves. *Perspectives of New Music* 28(1):58–78.

———. 1992. *Formalized music.* Revised edition. Stuyvesant: Pendragon Press.

Yoo, L., and I. Fujinaga. 1999. A comparative latency study of hardware and software pitch-trackers. *Proc. Intl. Computer Music Conf.* 1999:36–39.

Young, G. 1989. *The sackbut blues: Hugh Le Caine, pioneer in electronic music*. Ottawa: National Museum of Science and Technology.

Yu, C. 1996. *Computer generated music composition*. M. Eng. thesis, M.I.T.

Zadeh, L. 1965. Fuzzy sets. *Information and Control* 8:338–353.

———. 1975. The calculus of fuzzy restrictions. *Fuzzy sets and applications to cognitive and decision making processes*, ed. L. Zadeh et al., 1–39. New York: Academic Press.

Zendrum. 2001. *Zendrum Tech/MIDI*. <http://www.zendrum.com/techmidi.html>

Appendix

This appendix contains the full score to COALS. The first movement is called *Camera obscura*, the second *Carta terrae incognitae*.

Pitch and loudness of the eviolin should approximate that of a conventional violin. Timbre of the eviolin changes as the instrument moves horizontally in a space about one meter square. From the eviolinist's point of view, the spectrum is thin at left and rich at right, dull towards the back and bright at front. Positions in this square are notated in the score as dots in a square: the top left corner of the notated square corresponds to the front left corner of the horizontal square.

The other instruments also have certain requirements. The flute needs a B foot, and the trombone uses a plunger mute. The bass guitar should have both a bridge pickup and a neck pickup, which produce treble and bass tone respectively; either a switch or a "tone pot" knob can be used to select treble, intermediate *modo ordinario*, or bass tone. The B-flat clarinet and violoncello have no special requirements.

A few remarks are in order for notations for all instruments. Almost every note head has an accidental; exceptions are made for clear cases like tied notes. (Accidentals do, however, remain in force until the next bar line.) The smaller note heads of grace notes should not misleadingly cause them to be automatically played softer than non-grace notes. Flurries of grace notes should not be rushed at the expense of clarity: each note should be heard. Glissandi are notated as diagonal lines. As a rule they should be fast and not emphasized dynamically.

Vita

Camille Martin Anthony Goudeseune was born in 1966 in Eindhoven, The Netherlands. He received a Bachelor of Mathematics, honors co-op double major in pure mathematics and computer science, from the University of Waterloo (Canada) in 1989. Thereafter he worked in Redmond, Washington for two years as one of the software developers of Microsoft Word. He concurrently studied piano with Lora Verkhovsky and presented several recitals in Seattle.

From Seattle he bicycled to the University of Illinois at Urbana-Champaign to study music, first obtaining a Master of Music in piano performance and literature studying with Andrew De Grado. During this time he was principal developer of the VSS Virtual Sound Server while working at the NCSA Audio Group. He also co-founded a multimedia production company in Urbana and there produced *The Ring Disc*, a critically acclaimed CD-ROM of Wagner's opera cycle *Der Ring des Nibelungen*. Thereafter he completed his Doctor of Musical Arts in Composition with a minor in Computer Science.

He has studied composition with Erik Lund, Guy Garnett, and Herbert Brün. He composes for forces including tape, live electronics, chamber ensembles, and choirs.

He conducts research in real-time audio synthesis and augmented reality at the Integrated Systems Laboratory, Beckman Institute for Advanced Science and Technology in Urbana.