Control parameters for musical instruments: a foundation for new mappings of gesture to sound

DANIEL J. LEVITIN,† STEPHEN McADAMS‡ and ROBERT L. ADAMS§

†Departments of Psychology and Music Theory, and Centre for Interdisciplinary Research in Music, Media and Technology (CIRMMT), McGill University, Montreal, Canada
E-mail: levitin@psych.mcgill.ca
‡Institut de Recherche et Coordination Acoustique/Musique (IRCAM-CNRS), Paris, France
§University of California at Davis, USA

1. INTRODUCTION

In this paper we describe a new way of thinking about musical tones, specifically in the context of how features of a sound might be controlled by computer musicians, and how those features might be most appropriately mapped onto musical controllers. Our approach is the consequence of one bias that we should reveal at the outset: we believe that electronically controlled (and this includes computer-controlled) musical instruments need to be emancipated from the keyboard metaphor; although piano-like keyboards are convenient and familiar, they limit the musician's expressiveness (Mathews 1991, Vertegaal and Eaglestone 1996, Paradiso 1997, Levitin and Adams 1998). This is especially true in the domain of computer music, in which timbres can be created that go far beyond the physical constraints of traditional acoustic instruments. Key to our approach are the following three ideas:

(1) Suitable mappings must be found between a musician's gesture and the control of various aspects of a musical tone (Cadoz, Luciani and Florens 1984, Wanderley 2001b).
(2) Gestures are motions of the body that contain information (Kurtenback and Hulteen 1990).
(3) Mappings are best when they are intuitive, and when they afford the maximum degree of expression with minimal cognitive load (e.g. Keele 1973, Mulder, Fels and Mase 1997).

We hope that thinking about musical events in the manner we describe here will inspire designers, engineers and musicians to rethink the design of musical controllers in a way that will increase creativity and expression in synthesis and computer musical performances (c.f. Wanderley 2001a).

Through our diagrammatic interpretation of musical event control (figures 1–8), we will describe a new scheme for characterising musical control space. We first introduce a way of segmenting a single musical tone into five components that are conceptually distinct from a control standpoint. We then elaborate those five components by introducing new terms and conceptual structures based on control parameters. We believe that this parsing of a single musical tone into separate components has a valid and empirical basis, and that each component makes separate demands on the designer of new instrument control structures (and the player of new interfaces). We conclude with brief examples of how this conceptual structure might be used for mapping gestures to instrumental control.

The existing real musical instruments we know are the product of an evolutionary process – natural selection – only those which successfully balance expressiveness, clarity of control alternatives, and pleasing timbres have survived. In designing electronic controllers we cannot go too far wrong by emulating them. But we can consider this as just a starting point. Electronic instruments, less constrained by physics and acoustics, invite ever more imaginative approaches to controller design (e.g. Smyth and Smith 2001, 2002).

We begin first with some preliminary definitions and background in section 2, and then describe the components of a musical event in detail. We conclude with suggestions for how gestures might be mapped onto the parts of the musical event we describe.

2. MUSICAL CONTROLLERS VS MUSICAL INSTRUMENTS

The term *musical instrument* is commonly used to refer to a device that allows one to produce a variety of musical sounds. In the past two decades, various forms of computer-based synthesis techniques such as additive synthesis (Chamberlin 1976, DiGiugno 1976), physical modelling (Hiller and Ruiz 1971a, b, McIntyre, Schumacher and Woodhouse 1983), FM (Chowning 1973), wave-guide (Smith 1992), and scanned synthesis (Verplank, Mathews and Shaw 2000) have allowed for the computer to create musical sounds and to play back those sounds through loudspeakers. This has given rise to a need for an entirely new type of hardware device, the *musical controller* (or *input device*), a piece of equipment that the player uses to control how the sounds stored in (or created in real time by) the computer will be released (Roads and Strawn 1985, Mathews and Pierce 1989). This separation of the sound-generating source (in the computer) and the sound control source (typically in the hands and/or mouth of the player) is only a recent notion; in traditional acoustic instruments these functions are integrated (Mathews 1963, Hunt, Wanderley and Kirk 2000, Wanderley 2001a). In the clarinet, to take a concrete example, the musician blows into and fingers the very same device that is generating the sounds; the distinction between the sound generator and the controller is hard to see.

This functional distinction between sound source and controller is often blurred in contemporary electronic instruments as well. This is because for the last two decades the most common controller of synthesised sounds is modelled in appearance after the piano keyboard (Roads 1996). An actual piano keyboard is a complex mechanical device, and pressing down on a key initiates a chain of unseen events, including the movements of jacks, levers, hammers, and so on. In contrast, the synthesizer keyboard resembles the piano keyboard only superficially – after the initial key press, no such chain of mechanical events occurs, rather the chain of events is electronic. The synthesizer keyboard typically has one or more electronic sensors at the bottom of the key to sense various attributes of the key press (such as when the key was pressed, how fast or hard it was pressed, and when it was released; some keyboards also track 'aftertouch', pressure applied to a key when it is at the bottom of its throw). Although many early synthesizers were built as integrated units with a keyboard and computer synthesis module inside a single casing, this was simply a matter of convenience and a marketing decision. Functionally and conceptually, the portion of the instrument that created the sound (the computer) could be housed separately from the 'input device', the keyboard. Most modern synthesizers are in fact available as either an integrated unit (keyboard controller and computer in one) or as separate units.

The functional and physical separation of the sound source and the control of that source opens up new possibilities for alternative control devices (Wanderley 2001a). Although the electronic piano keyboard has served well as an input device (because it is familiar to many players), a number of other input devices are possible. In 1987, Yamaha introduced a wind controller (the WX7) that resembled a soprano saxophone, and allowed the player to control the sound output of a MIDI sound module by blowing and pressing keys (as of this writing, the current version is the WX5). Various other input devices have included electronic controllers made to resemble drum heads or guitars, and even typewriter keyboards have been used (Roads 1996). A separate class of controllers are only loosely (if at all) based on traditional acoustic instruments, and are designed to translate the musician's intuitive gestures more directly into sound manipulation (see, for example, Mathews 1991, Rich 1991, Boulanger and Mathews 1997, Mulder, Fels and Mase 1997, Marrin-Nakra 2000, Buchla 2002). Such alternative musical controllers are not only worth exploring for the expressive and control advantages they offer, but can reduce repetitive strain injuries that have become a serious problem for many performing musicians playing traditional instruments (Markison 1990).

Research in synthesis and computer-generated sound for the past decade has placed great emphasis on the development of better sound while comparatively little attention has been paid to the devices that would control that sound. In contrast, traditional acoustic instruments have evolved a great diversity of control structures (Hunt *et al.* 2000). In the computer era – in which we have access to an unprecedented range of sounds – there is no reason to remain shackled to the keyboard as the principal means of controlling that sound.

As Curtis Roads notes:

> Electronic input devices detach the control of sound from the need to power the sound; any one of dozens of input devices can control the same sound generator. This translates into musical flexibility. With electronic instruments, a single wind controller can create the low bass sounds as easily as the high soprano sounds. Creating extremely soft or loud sounds requires minimum effort since the control is electronic. Obviously, the detachment of sound control from sound production has a negative side – the reduction of the 'feel' associated with producing a certain kind of sound. (Roads 1996)

One hurdle to the easy development of new sound controllers is that the systematic study of how musical sound can be controlled has not received much scientific attention (Wanderley 2001a). The recent separation of sound generation from sound control has created the need for a new classification system based on *control parameters*, in order to facilitate the principled design of new instrument controllers. Previous classification systems emphasised either the physical origins of sound production (Sachs and Hornbostel 1914/1961), or the perceptual parameters of sound (Vertegaal and Eaglestone 1996). Here, we attempt to combine these two approaches and introduce an emphasis on what is important to know when one wants to *control* musical sounds.

Before describing our new classification system for musical control, we will briefly review some things that are known about traditional musical instrument control.

2.1. A brief review of traditional musical instruments

What is it that makes some musical instruments easier to play than others, or more enjoyable or more expressive than others? Many more homes in the U.S. seem to have pianos than flutes, even though flutes are lighter, less expensive, and novices who devote a year of study to each instrument achieve roughly equivalent levels of competence. Something about the piano draws people in. Aside from social considerations (the ownership of pianos has long been perceived as a status symbol regardless of one's inability to play one), perhaps there are properties of the instruments themselves that affect people's decisions – and their abilities – to play the instruments.

One of the factors affecting people's decisions about whether to take up a musical instrument (and if so which instrument) is the 'learning curve', or the amount of time it takes a novice to gain enough skill with the instrument that the experience of playing it is rewarding (Vertegaal and Eaglestone 1996). Any human-made device must strike the right balance between challenge, frustration and boredom: devices that are too simple tend not to provide rich experiences, and devices that are too complex alienate the user before their richness can be extracted from them (Levitin and Adams 1998, Wanderley and Orio (submitted)).

The machines with which we have the most rewarding relationships are those that allow us to interact with them through the development of a finely tuned skill. All of us have seen experts practising their craft and the delicate and exquisite skill they bring to the activity. Tennis players, race car drivers, jewellers, carpenters and musicians are all examples of people who use a tool for their profession, and the best of them use these tools in ways that are very different from the rest of us, ways that are the consequence of years of careful training. Most of us are able to use a tennis racket, to drive a car, to work with a jeweller's pliers, a carpenter's hammer, or to sit at a piano and press the keys. This is what contributes to the success and ubiquity of these technological artifacts, the fact that they are usable by most of us with a little training, and yet they are rich enough that one can spend a lifetime developing skill on them. Each of these tools strikes the right balance between ease of initial use (low frustration) and richness of continued use (ongoing challenge).

The learning curve for each musical instrument is related to the ways in which it is controlled. Two important factors are whether the instrument's controls are arranged in a logical fashion, and whether a given gesture on the instrument produces identical (or similar) sounds across multiple occasions. Apart from the various acoustic and physical factors that distinguish the instruments from each other, it would be useful to know which instruments share similar control structures. This knowledge could form an avenue into a deeper appreciation for how people learn and use instruments based on the part of the instrument they actually touch and manipulate, the 'controller'.

Before developing a classification scheme or taxonomy for musical control (the principal topic of this paper), it is useful for a moment to forget about the differences among individual musical instruments and just consider them as a group. One could simply make a list of the different ways that these tools let us control and manipulate sound. The list would include activities such as these:

- select a tone from among a 'menu' of tones,
- start a tone playing,
  - add a tone to an ongoing tone (in polyphonic instruments),
- stop a tone from playing,
  - take away a tone from an ongoing set of tones (in polyphonic instruments),
- create vibrato (small, rapidly oscillating shifts in pitch),
- create tremolo (small, rapidly oscillating shifts in volume),
- trill (alternate rapidly between two or more tones),
- glissando (slide to a new pitch, producing continuous pitch change in between),
- select a particular dynamic,
- change dynamics (louder or softer),
- select a particular timbre,
  - change timbre,
  - create various timbral or spectral effects specific to the instrument (e.g. 'growling' on the saxophone; feedback or harmonic shifts on the electric guitar; pedalling on the piano; various embouchure changes).

Our next task is to generate a principled model that separates out groups of control structures into meaningful clusters. An initial important distinction we can make divides the family of musical instruments into two groups, the polyphonic versus the monophonic instruments.

2.2. Polyphony vs monophony: factors affecting cognitive load

Polyphonic instruments allow for the player to create more than one tone at a time; monophonic instruments allow only one tone to sound at a time. Polyphonic instruments include pianos, guitars, violins, accordions and xylophones. Monophonic instruments include clarinets, trumpets and most drums. (Of course, there are grey areas. For example, some wind players can create two or more tones at once through special techniques such as multiphonics, but these are rarely used and at any rate, these different tones cannot easily be controlled independently of one another.)

In the previous section, we listed some of the ways that a musical tone can be manipulated, such as vibrato, tremolo, dynamics and so on. On polyphonic instruments the player could theoretically choose to manipulate these parameters on more than one tone at a time, or to manipulate them for one or more tones as they are added to existing tones. In the history of musical instrument development across the last several thousand years, no known polyphonic instrument has ever allowed for all of the possible control and manipulations outlined above. The reason for this probably has to do with both *physical limitations* of the musician, and *information processing limitations* of the human brain, either due to limits of *cognitive loading or to coding limitations* – that is, the inability to store and recall too large a number of settings, positions, or sounds (Fitts 1954, Miller 1956, Fitts and Posner 1967, Cook 2001).

As an example of physical limitations, consider what would happen if we wanted to build a trombone that could play two tones at once. The way the trombone is currently played, the musician moves the slide with her right hand and grips the instrument (in order to stabilise it) with the left hand. We could provide a stand to stabilise the instrument, freeing up the left hand to operate a second slide. But different tones require adjustments in lip pressure and embouchure, so our first hurdle arises because it is unlikely that our player would be able to play into two mouthpieces at once, or to create two different tones using one mouthpiece. A second hurdle arises when we realise that our duo-bone will require twice as much air to play.

Suppose that we are successful at training a musician to play out of both sides of her mouth, making the precise lip adjustments required to produce two different tones, and that further, she has learned to produce an adequate amount of wind, and she can control both slides because the duo-bone is sitting on some sort of stand. Having crossed these formidable physical hurdles, the mental effort required to master the control of all these parameters would undoubtedly defeat most musicians – there are just too many things to keep track of (Cook 2001). This is an example of the limit of the brain's information processing capacity (c.f. Miller 1956, Keele 1973).

Another example involves the control of the piano. If we ignore for the moment ancillary controls such as pedalling, the pianist has only four decisions to make for a given tone to be produced: *which* key to press, *when* to press it, *how fast* to press it, and *when to release it*. Of course the production of sequences of tones and the adjustment of their phrasing is clearly more complicated and requires many more decisions. But the striking fact is that the entire rich expanse of tonal colours that a beautiful grand piano can create results from controlling just these four parameters. However, there are some things pianos do not do. Because pitch selection is constrained to be *discrete* (the keys on the piano only match to the focal tones of the scale), pianos are incapable of playing any of the musical sounds between scale tones.

In contrast, the player of a guitar, kettle drum or violin can *slide* from C to C♯, and can even control how long he wants to take to get from one tone to the other, and what intermediate tones he wishes to emphasise. The piano player, however, is stuck with C and C♯. As a consequence, pianos cannot produce vibrato, glissandi, bending effects, and cannot use 'expressive intonation' (the intentional flatting or sharping of tone for emotional effect).

It is probably a good thing that pianos do not allow these types of control due to limitations of information processing capacity. Whereas the piano allows for the simultaneous playing of many different tones (constrained only by the number of fingers the player has), if the player had to control the vibrato level, or for example, the placement of his finger within the key in order to choose the precise intonation of that tone, this would present significant physical and mental hurdles. No one has ever built a grand piano that allowed these kinds of controls, not because it would be difficult to build one, but presumably because it would be so difficult to play one.

3. OVERVIEW OF THE CONTROL OF A MUSICAL INSTRUMENT

Our goal in the remainder of this paper is to analyse the various ways of controlling monophonic musical instruments. Our hope is that thinking about control in this way will reveal certain principles that designers can use in creating new instrument controllers. We believe that it is beneficial to emancipate musicians from the dominant 'piano metaphor' that, while ubiquitous in synthesis, is an impoverished and limiting constraint. In the model that follows, we are leaving out the perceptual dimension, except for occasional asides.

As figure 1 indicates, musical events unfold over time, and they can be thought of as occupying different regions along the temporal continuum. A single cycle of a sound wave lasts a very short time, usually measured in milliseconds: one cycle of a 120 Hz tone, for example, lasts approximately 8.3 milliseconds, and that of a 3,500 Hz tone lasts only 285 microseconds. Groups of cycles comprise a tone, which might be measured in fractions or whole seconds. Groups of tones form musical phrases (typically measured in tens of seconds), and groups of phrases form a composition (typically measured in minutes or tens of minutes). The type of analysis we develop here could properly be applied to various musical events along this timeline, but our goal in this article is to systematically elaborate what it is we know about a single musical tone (more than a single sound wave cycle, but less than a phrase).

Because musical tones unfold over time, acousticians have found it useful to divide up a given tone into three portions they refer to as the *attack, steady state and decay* (or *release*) of a tone (Pierce 1983, Taylor 1992). In simple terms, these can be thought of as the beginning, middle and end of a tone's development over time (bottom half of figure 1). The *attack* of a tone refers to that portion of the sound that is a direct and immediate consequence of whatever mechanism was used to *begin* the musical tone. Thus, striking a surface, blowing into

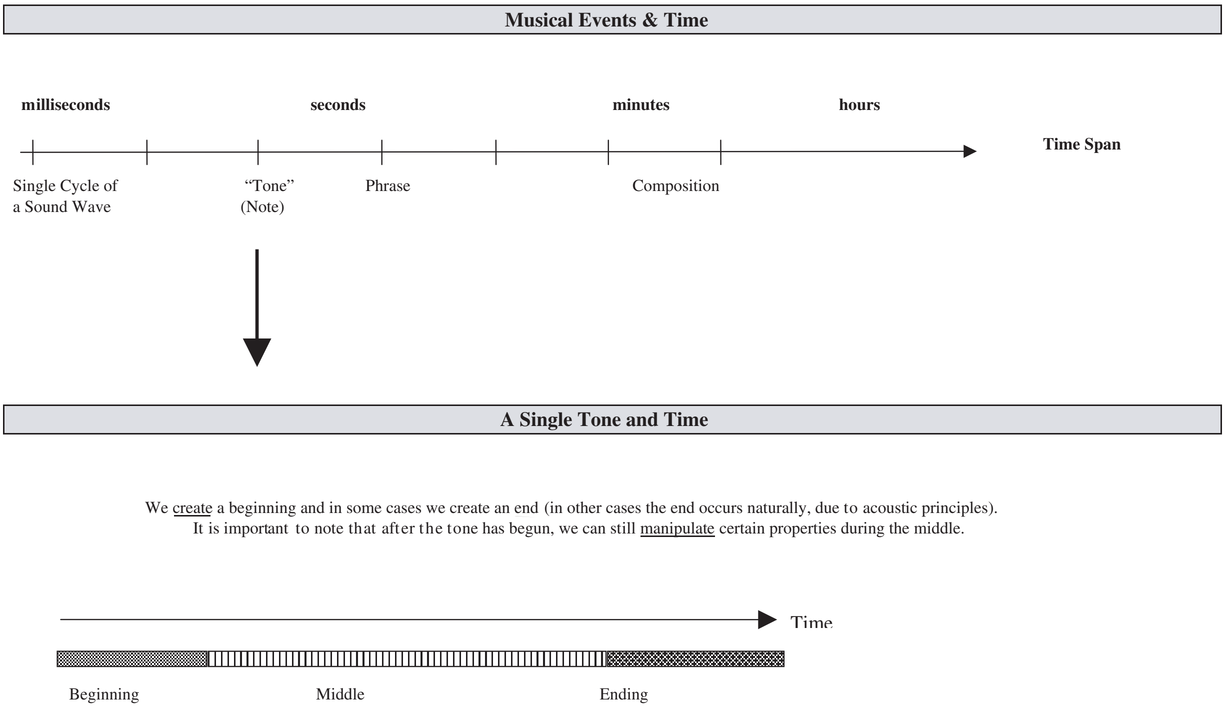


Figure 1. Top half: musical events unfold over time, from milliseconds to hours. Bottom half: a single musical tone is made up of three components, distinct in terms of their control and perceptual attributes.

a mouthpiece, bowing or plucking a string are all examples of an *excitatory gesture* that introduces energy into a system, sets up some form of vibration, and results in sound being created from the instrument (Cadoz, Luciani and Florens 1984).

The reason that acousticians distinguish the *attack* from subsequent portions of the tone is that in many cases the attack has specific acoustic properties (spectral, temporal and spectro-temporal) that are primarily the result of the initial contact with the instrument, and these specific acoustic properties are thus very different in character from the remainder of the tone; these properties are often called 'transients' and contain non-periodic and 'noisy' sounds. The boundary between the 'attack' and the 'steady-state' portions of the tone is fuzzy, but the 'steady state' (or 'middle') is the portion of the sound in which the primary resonator is the dominant contributor to the sound heard, and is so called because in most instruments that portion of the sound is periodic.

Many of the perceptual cues we use to distinguish one instrument from another occur during this 'attack' phase of the tone (Stumpf 1926, Schaeffer 1966, Risset and Mathews 1969), and its steady-state portion can be somewhat difficult to characterise uniquely, unless it has particularly salient and distinctive spectral characteristics like the weak even-numbered harmonics in a clarinet tone, for example. Thus, if we electronically edit out the attack of a piano tone and an oboe tone, the instruments are surprisingly difficult to tell apart (Saldanha and Corso 1964). These phenomena were known to Stumpf as early as 1910, and later elaborated by Pierre Schaeffer (1966) in the famous 'cut bell' experiments.

Charles Taylor (1992) performs a classroom demonstration to make this concept even clearer. He has three musicians playing a clarinet, a flute and a French horn, play from behind a screen. The players all begin a tone at different times, and it is easy for most students to determine the order of the instruments entering. However, Taylor then has the players all start together but drop out one by one. In this case it is very difficult for students to tell the order. This is because so much timbral information is contained in the 'attack' portion of the tone, and relatively less distinguishing information is contained in the 'steady-state' or 'decay' portions of the tone. In addition, we have neural onset detectors, but no such offset detectors (Kolb and Whishaw 1990). When the instruments drop out, because it is the relatively undistinguishable 'steady-state' portion of the ongoing instrument that is still heard, it is difficult to determine with any certainty what that instrument is (Cook 1999).

The terms 'attack', 'steady-state' and 'decay' are useful from an acoustical point of view. But from a control perspective, they are inadequate and confusing. For one thing, the so-called 'steady-state' of a tone (implying acoustic stability) is the portion of the tone when the player typically has the *most* control over the sound – thus it is the portion when the tone is potentially the *least* stable from a control perspective. Furthermore, in everyday use, the word 'attack' suggests an explosive

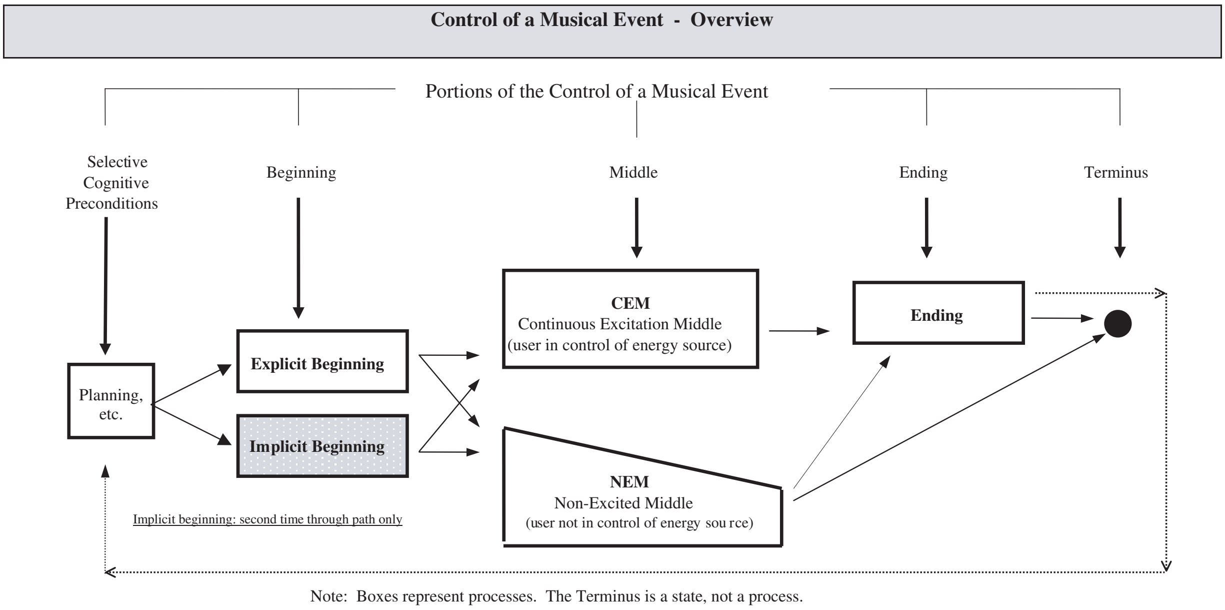


Figure 2

or high-impact beginning, but in fact many tones are begun without this sort of explicit event, such as when a flute player – while continuing to blow into the instrument – simply lifts her finger off of one key to change pitches and create a new tone. Even as used by acousticians, the word 'attack' suggests some sort of sudden introduction of new energy to the system. In the case of a change of fingering, although transient components may exist, the effect is qualitatively different from an attack.

Since our goal is to define a system that will facilitate the design of musical controllers, we need to reconsider the way we think about these portions of a musical event. Given that the words 'attack', 'steady-state' and 'decay' are encumbered with meanings, and because they are used in a reserved way by acousticians, we will instead refer to the three portions of the control of a tone as the *beginning, middle* and *ending* (see figure 2). In addition to these three portions of controlling a tone, there are some preliminary and concluding elements required to make our model complete (labelled 'selective cognitive preconditions' and the 'Terminus' in figure 2) and we will discuss those shortly. Figure 2 displays a *logical path* of the five stages we believe constitute the control anatomy of a musical tone. We are not talking about the anatomy of a tone *per se*, but of the control of that tone.

The first thing our model of musical control needs to account for are the cognitive processes involved in preparing to play a tone (the first box in figure 2). Prior to physically beginning a tone the musician has to make a variety of decisions such as what tone to play, how loud to play it, when to play it, and so on. These *psychological factors* are part of the *planning process* that normally must occur before the musician makes any sound come out of the instrument (what Cadoz, Luciani and Florens (1984) termed the 'preparatory phase'), and with practice they may become automatic or subconscious. These collective planning processes are what we mean by the *preconditions* to beginning a tone. They include not just strictly mental preparation such as deciding what to do, but physical preparations, such as raising one's arms in preparation to striking a drum, drawing in a breath in preparation to blowing a trumpet, drawing back the violin bow, etc. Since the musician typically must *select* from an inventory of choices available to her, we call these *selective preconditions*. In other words, the number and types of actions the player can choose are constrained by the design of the instrument and are not unlimited (we are speaking here of traditional acoustic instruments).

Once the selective preconditions have been set, the player initiates the musical tone and the *beginning* of the tone occurs. It is important to be clear that we are distinguishing between two types of *beginnings* depending on whether another tone was already playing when the current one begins. That is, there is a conceptual difference – and an important control difference – between the case in which our flute player creates a new tone by starting to blow (what we call in our model an *Explicit Beginning*), and the case in which the player is already playing a tone and already blowing, and creates a new tone by changing fingerings or tonguing (what we call an *Implicit Beginning*). We elaborate on this distinction in section 3 below, but the essential point is that in the former case energy is being added to the instrument (in this example the energy of course is air), and in the latter case, energy is being continuously applied to the instrument and a new tone is created without an explicit or new excitation to the instrument. Thus, *Implicit Beginnings* cannot occur on the first time through this logical path model, only on subsequent times, because a tone has to be already playing.

Continuing through our anatomy of the control of a tone, the *beginning* of the tone is followed by the *middle* of the tone, and here we make the distinction between two types of middles, depending on whether or not the user is in control of the energy source during this middle portion of the event. In one case, the user is still applying energy to the instrument during the middle in the same manner as she was during the beginning. The fact that she is still applying energy allows her to alter certain parameters of the sound differently than if she were not. We call this type of middle a *Continuous Excitation Middle* (CEM) such as is created by a bowed violin or a saxophone (forced-vibration instruments). This is in contrast to those instruments in which the user is no longer supplying energy to the instrument during the middle, and hence has more limited or even no control over the sound. We call this second type of middle a *Non-Excited Middle* (NEM) and it includes those middles that are created by instruments such as cymbals, drums and pianos (impulsive or percussive instruments). (The two types of middles will be further elaborated in section 4 below.) Note that a Non-Excited Middle tends to be characterised by a reduction in intensity (because by definition there is not a continuous supply of energy) and we indicate this by the downward sloping box in figure 2.

In the flow of the logical path in figure 2, either type of middle (CEM or NEM) can follow either type of beginning (Explicit or Implicit). However, the type of middle constrains the choices available for the next stage of the model. Following Continuous Excitation Middles (CEMs) the player must make an explicit decision to end the tone (by withdrawing energy) and the point in time where this decision is implemented marks the next stage, the *Explicit Ending*. (The ending is not instantaneous, but is a process; what we mean by this requires some elaboration which we detail in section 5 below.) At some point, the instrument actually stops producing sound and we call this point the *Terminus*. Note that the Terminus is not a process stretched out over time as are the previous modules of figure 2, but rather it is a single point in time; as such, the Terminus represents arrival at a *state* not a *process*. Note also that for those instruments that have a Continuous Excitation Middle, the Terminus must follow an Explicit Ending (a decision to withdraw the supply of energy). However, for those instruments that have a Non-Excited Middle, the player has a choice. He can either damp the instrument (a form of *Explicit Ending* that then leads to a *Terminus*) or he can just allow the sound to decay on its own (eventually leading to a Terminus without the intermediary *Explicit Ending* step).

We have gone to some effort to clearly distinguish the Ending and the Terminus. Why? There are two reasons we cannot include the Terminus in the ‘Ending’ process. One is that as we just noted, some musical events terminate directly after the Non-Excited Middle stage (refer to figure 2). The other reason is that the ‘Explicit Ending’ process by definition covers a time period in which sound is being produced. We need a definite point that denotes when the musical event has terminated, the structural equivalent in this model of a dummy variable to act as a place-holder for the state of ‘sound no longer playing’, and we call this dummy variable the *Terminus*.

4. CONTROLLING THE BEGINNING OF A MUSICAL EVENT

Figure 3 elaborates various aspects of the control of the *beginning* of a musical event. Unlike figure 2 which is intended as a logical path diagram with *time* running along the *x*-axis, Figures 3–5 are intended to convey different conceptual aspects of a single portion of control over a musical event. That is, figure 3 provides an expanded definition and conceptual elaboration of ‘The Beginning’, and the *x*-axis represents various ways of thinking about the beginning under the headings ‘Type of Attack’, ‘Perceptual Cue’, ‘Theoretical Description’ and ‘Gesture’. It might be helpful to read figure 3 from right to left: the inventory of Gestures shown at the right hand side of the figure represents things the musician *does*, and the various columns to the left represent consequences of that action.

The first way of thinking about controlling the Beginning of a musical tone we call the ‘Type of Attack’. As mentioned above, a Beginning can be either *explicit* or *implicit*, depending on whether it is a brand new event or a modification of a pre-existing event, respectively. These two types of beginnings are uniquely distinguished on three additional dimensions, the *perceptual cue* that the listener hears; the *theoretical description* of how the beginning came into being; and the physical *gesture* the player used to create the beginning.

The second way of thinking about the Beginning of a musical event is in terms of its perceptual qualities. Explicit Beginnings are always accompanied by a discontinuity along the sound intensity dimension. This is because by definition, an ‘Explicit Beginning’ is the term applied to a tone that is produced ‘from scratch’, so an Explicit Beginning of a tone is marked by a sudden new sound where there was no sound before. (And if this is not the first tone to be created in a musical phrase, it must come from the terminus state.) If we graph sound intensity versus time, the beginning of a tone thus appears as a discontinuity in the graph. While this may seem rather obvious, it forms an important contrast to the way in which *Implicit Beginnings* manifest themselves perceptually. By definition, an Implicit Beginning (in monophonic instruments) is the term applied to the

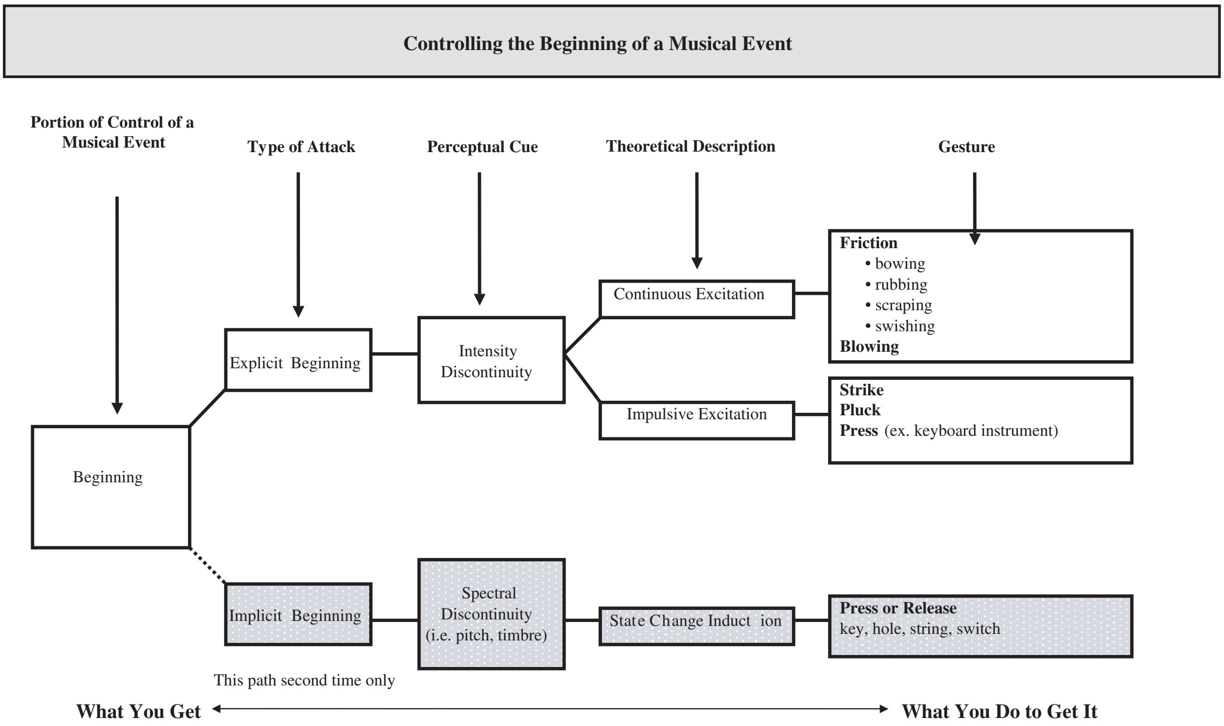


Figure 3

creation of a new musical tone when another tone was already ongoing. The way that we decide a new tone has started, as opposed to the previous tone just slowly evolving into a new one, is a perceptual/psychophysical question, and it is not always clear (Bregman and Rousseau 1991). But if there is a sharp spectral discontinuity, say in pitch or timbre, our perceptual system will tend to identify the portion of the ongoing sound that follows the discontinuity as a 'separate object'. (Thus in figure 3, for the case of an Explicit Beginning we represent 'Intensity Discontinuity', and for the case of an Implicit Beginning we represent 'Spectral Discontinuity' under the heading 'Perceptual Cue'.) This idea is consistent with the Gestalt Principle of Good Continuation, which specifies that abrupt changes along some mapped dimension typically indicate that there are two objects or events in the physical world, not one (Wertheimer 1923/1938, Bregman 1990).

The third way of thinking about controlling a Beginning we call the 'Theoretical Description'. In any Explicit Beginning, the musician introduces energy to the instrument. This can be done in two ways: (i) the musician introduces energy in an impulsive fashion for a brief period, or (ii) the musician introduces energy in a continuous fashion. There is a discrete set of *gestures* that accompany these two theoretical descriptions of the manner in which the tone was begun. For 'impulsive excitations', the musician might strike something, such as a drum or a xylophone, either with the hand or a mallet or stick. Other impulsive excitations are all basically variations of the 'striking something' category, and include plucking and pressing a key. Thus, in the gesture column of figure 3, we inventory various types of movement, including these 'ballistic' or 'striking' movements. Note that except for percussion, these are rarely used in controlling musical instruments (Bonnet, Guiard, Requin and Semjen 1994, Wanderley 2001).

For 'continuous excitations', the musician either blows into the instrument, or excites the resonator via friction, such as by bowing, rubbing, scraping, or activates a switch. Squeezing an accordion or bagpipe bellows is equivalent in this sense to blowing into the instrument. We note that the description here is adequate for traditional acoustic/mechanical instruments, but inadequate for electronic instruments. For example, a synthesizer player can create Continuous Excitation simply by holding down a key.

Thus, we can follow the diagram for the Explicit Beginning and see that it always arises from a perceived intensity discontinuity, and that there are two ways in which that intensity discontinuity could have been created, either by one of the set of gestures that cause 'continuous excitation' or by one of the set of gestures that cause 'impulsive excitation'.

In the case of an Implicit Beginning, there is no perception of an intensity discontinuity (by definition), so the identification of a 'new musical event' has to be made on the basis of some sort of spectral discontinuity, such as an abrupt change in pitch or timbre. When we discuss the 'Middle' of a tone in the next section, we will discuss the types of modifications that are made to an ongoing tone, such as vibrato, tremolo, changes in loudness, and so on. It is clear that certain spectral properties of a tone can change and the listener will identify the sound as comprising a single musical event, whereas in other cases a sufficiently large change in spectral properties of a tone cause the listener to believe that the old tone is 'gone' and a new one has arisen to take its place (Bregman 1990). The thresholds for such identifications will vary from listener to listener, and are also dependent on musical context. But for our purposes, perhaps it is sufficient to acknowledge that there are clear cases at the ends of the continuum of spectral change, and the matter of fixing the precise location of the boundary need not trouble us for now.

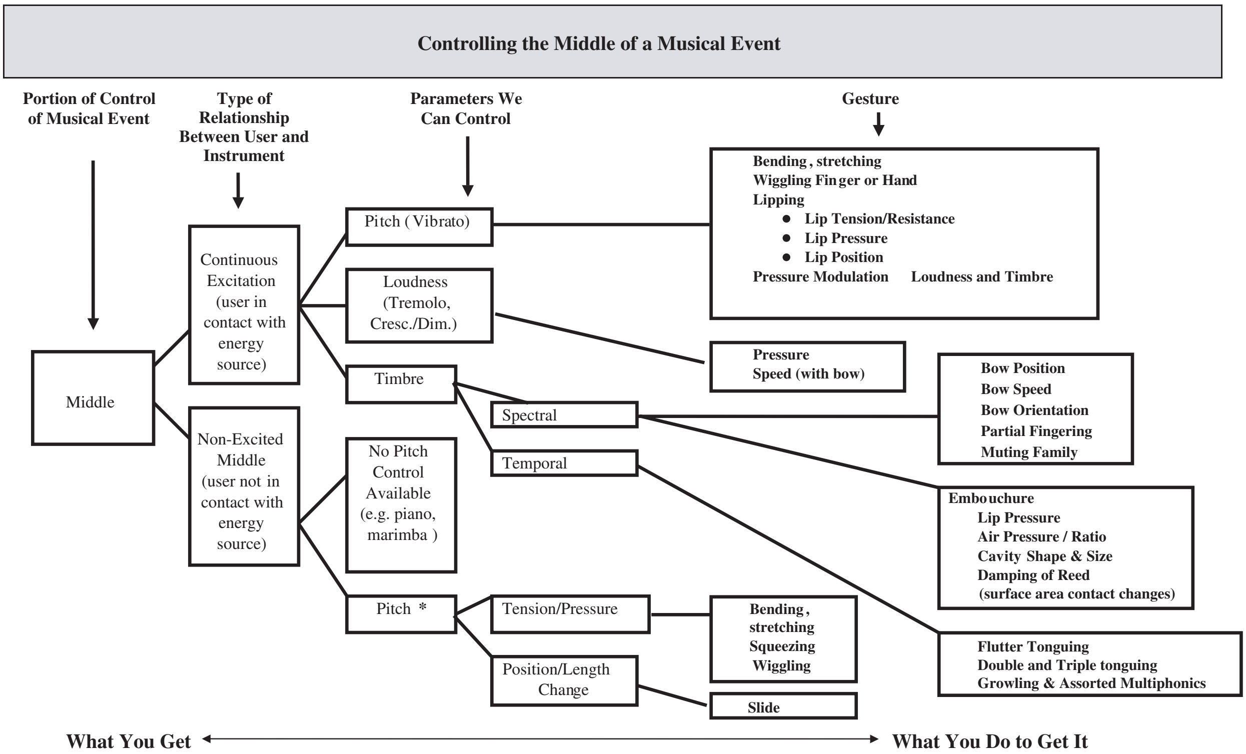
One fuzzy case that should be mentioned is that of a *glissando*, which we consider a type of Implicit Beginning. How do listeners interpret the case of a tone that glides from one pitch to another? The logical possibilities are that it could be interpreted as a single, continuous tone changing in pitch; as two tones (one at the initial and one at the final pitch); or as some large – perhaps infinite – number of tones corresponding to every pitch between the first and last tone. Common experience and parsimony dictate that most glissando events are perceived as two tones, carrying the initial and the final pitch, provided that the initial and final pitch last long enough to establish themselves perceptually (we intentionally avoid a discussion of what 'long enough' means, and leave that to psychophysicists). This is reflected in musical notation, in which a glissando is written to show the first pitch and the last pitch with a line in between.

But what exactly is the brain doing during the glissando? How does it 'decide' when the second tone begins if it does not already know what it is going to be? That is, we need to account for the phenomenological experience listeners have that as soon as the top tone of an upward glissando (or the bottom tone of a downward glissando) is reached, they know that the glissando is over and the tone they are hearing is actually a *new tone* (one of our 'Implicit Beginnings'). The computational perceptual problem is that as each new pitch is reached in the glissando, it could be the pitch that the instrument ends up 'staying on' and the brain would need to know *that* in order to identify it as the beginning of the new object. This is analogous to the so-called binding problem in visual perception, specifically as it regards the phi phenomenon or apparent motion (Dawson 1991, Dennett 1991, Palmer 2000). The answer is that the brain does not know *at the time of occurrence* of the glissando. Once the glissando is over, the brain *retrospectively* assigns the label 'new object' to the final tone of the glissando. It all happens so fast that we have the phenomenological experience of apparently knowing that the glissando has ended just as its final tone begins! However, it seems clear that the brain must come to the conclusion that the pitch has stopped changing long enough for the new note to be recognised as 'established', and that the brain does not do this until *after* the glissando has finished. Working in support of this illusion are expectations and memory; when we know a musical piece, our brain anticipates the arrival at the final point of a glissando.

Returning to our diagram and the overview of Implicit Beginnings, we see that these arise when the perceptual system (by some means) identifies that a new tone has begun, based on the *perceptual cue* of a *Spectral Discontinuity*. This is accomplished by a discrete set of gestures made by the musician, such as pressing or releasing a key, string or switch, or the covering or uncovering of a hole. At the level of theoretical description, these gestures cause a spectral discontinuity by inducing a state change of some sort in the sound; thus we call this type of control of the sound a *State Change Induction*.

One key point about Implicit Beginnings is that generally *they can only result from a continuous excitation gesture*. That is, those state changes that result in an Implicit Beginning must follow Continuous Excitation Middles (CEMs) in our model, not Non-Excited Middles (NEMs). This is because CEMs create the auditory scenario within which it makes sense to talk about changing a tone. In the case of NEMs, the tone is decaying in intensity, the user is not in direct contact with the energy source (more about this in the next section), and so generally speaking we cannot properly talk about one tone emerging implicitly from out of another. Note also that in most cases, even though the musician is applying energy *continuously* to the instrument prior to (and during) a State Change Induction, the actual *gesture* that gives rise to the perception of a new tone (through spectral discontinuity) is itself usually *discrete*, not continuous, as in pressing or releasing keys, holes, strings and switches. One exception to this is the Implicit Beginning that can be created on variable pitched instruments through glissandi (such as bending a string or sliding the trombone slide). And as we stated above, a glissando could be considered a type of Implicit Beginning, due to the fact that (by definition) it describes something that happens when a tone is already ongoing. Another exception occurs when a guitarist (or other string instrumentalist) performs 'hammer-ons and pull-offs'. Although generally speaking the plucking of a guitar string is considered to occur as the result of an *impulsive excitation* from the right hand/picking hand (not a *continuous excitation*), the guitarist can hammer on or pull off a left-hand finger which effectively introduces new energy into the system, but in a way that sounds more like the adding of a finger in a wind instrument than the creation of a new tone. This hybrid case is an exception to the rule outlined above.

At the bottom of figure 3 we remark that descriptions in the far right-hand column, the 'gestures' are 'things that you do'. Descriptions on the left hand side of the chart are 'what you get' as a result of those things.



* Only if the player can contact some part of the tone generator during the Non-Excited Middle.

Figure 4

5. CONTROLLING THE MIDDLE OF A MUSICAL EVENT

After the tone has started, the ways in which it can be controlled are determined by whether the user/musician remains in contact with the energy source (figure 4). In other words, when the musician is using one of the ‘continuous excitation gestures’ such as blowing, rubbing or bowing, the user is in effect providing continuous excitation to the primary resonator. Contrarily, if the musician has used one of the ‘impulsive excitation gestures’ such as striking or plucking, the user is no longer in contact with the energy source, and the sound output of the primary resonator begins to decay immediately after the initial impulse.

A brief diversion to discuss ‘primary’ versus ‘secondary’ resonators. The primary resonator should not be confused with any secondary or tertiary resonators. The primary resonator is that part of the instrument (or physical model in the case of synthesis) to which energy is introduced directly. To elaborate the distinction between primary and secondary resonators, the bars on a xylophone are the primary resonators, and the hollow tubes below them are the secondary resonators. That is, once a vibration in a bar has been initiated (the user contacts the bar, not the tube) the energy is passed acoustically into the tubes which serve to filter and radiate the sound of the primary resonator.

As another example, on the chordophones (the term used in the Sachs and Hornbostel (1914/1961) classification system) such as the violin, guitar or piano, the string is the primary resonator. The secondary resonator takes energy from the primary resonator and filters and radiates it. Thus the body of the violin or guitar and the soundboard of the piano are secondary resonators. Accordingly, in the case of a bowed violin, the user is in continuous contact with the primary resonator. In pizzicato violin, strummed or plucked guitar, and piano, the user is *not* in continuous contact with the primary resonator. In some cases, the user may in fact be in continuous contact with a secondary resonator, but secondary resonators do not typically allow any great degree of control over the sound quality.

In wind instruments, the primary resonator is considered by some acousticians to be the mouthpiece (including the reed, if any), and the body of the instrument serves as a secondary resonator, filtering and radiating the vibrations of the mouthpiece. In membranophones (such as drums with stretched skins), the membrane is the primary resonator and the drum body is the secondary resonator.

5.1. Continuous excitation during the middle of a tone

Returning to the issue of whether an instrument is being continuously excited or is in a state of decay, this distinction affects the types of parameters the player can *control* during the *Middle* portion of the tone (see the heading 'Parameters We Can Control' in figure 4). In continuous excitation cases, the musician can manipulate the energy source itself (since energy is being continuously provided) to introduce various modifications to the sound. For example, some instruments allow for the ongoing *control of pitch* (typically manifested as vibrato or expressive bending of the pitch) and an inventory of the gestures commonly used to do this is given in the far right-hand column of figure 4:

- bending or stretching (as in a guitar string or the head of a kettle drum),
- wiggling finger or hand (as on the fingerboard of a violin),
- lipping (as on wind instruments), and
- pressure modulation (as on wind instruments).

We can further subdivide the category of 'lipping' into changes in lip tension or resistance, changes in lip pressure, or changes in lip position, all embouchurial gestures that cause changes in pitch. Note that pressure modulation proportionally affects loudness and timbral changes. Moving downwards in figure 4, under the heading 'Parameters We Can Control', we see that some instruments allow for the ongoing *control of loudness*, and this is usually implemented by a gesture of changing *pressure* or changing *speed* in the musician's contact with the energy source.

Finally, some instruments allow for the ongoing *control of timbre*. Timbral changes can be divided up into two types, *spectral* and *temporal*. Spectral changes are commonly controlled by:

- bow position,
- bow speed,
- bow orientation,
- partial fingering,
- various kinds of muting, and
- embouchure changes.

In turn, an inventory of embouchure changes includes lip pressure, air pressure, noise-to-tone ratio (breathiness), changes in the mouth cavity shape and size, and changes in the way the mouth and tongue contact the surface of a reed (various forms of reed damping). Temporal changes are controlled by:

- flutter tonguing, double- and triple-tonguing,
- growling, and assorted gestures that create multiphonics.

To recap, the above description applies to Continuous Excitation Middles (CEMs) because all of the gestural modulations mentioned require some modification of the primary resonator, and hence the user must be in contact with it.

5.2. Non-excited middles

The above is not meant to imply that the user has no control over the sound of the instrument in the case of a Non-Excited Middle (NEM), only that these cases involve a different conceptual structure. The majority of instruments that the NEM comprises do not allow pitch control during this decay portion of the sound event (e.g. piano, marimba, plucked guitar, snare drum, cymbals). There are a few rare cases in which the musician *can* control pitch during such a decay, and this is accomplished by the musician having some contact with the *tone generator* during the Middle, but by definition there is no contact with the *energy source* during an NEM because the energy source was impulsive and so is no longer present. An example is the kettle drum. After the membrane is struck (the membrane is the primary resonator) the musician can stretch the membrane using a pedal and thus affect pitch. In general, the *gestures* that accomplish changes in tension or pressure are:

- bend,
- stretch,
- squeeze,
- wiggle, and
- slide.

The only other way in which pitch can be changed in an NEM is through a position or length change. This is the case, for example, with a decaying tone on the guitar during which the player bends the string, or bends the neck to achieve pitch changes.

Finally, some timbral changes are possible in NEM instruments, but these are usually only very minor and not of practical importance. For example, after a chord on the guitar is strummed, the musician can hold the instrument up and shake it, causing the sound waves inside the body to propagate somewhat differently, but the effect is acoustically quite subtle.

6. CONTROLLING THE END OF A MUSICAL EVENT

Next we need to make a distinction between whether the ongoing tone is going to end in silence or whether the musician will go on to an Implicit Beginning (figure 5). For the case when the tone is going to end in silence, we define an 'Explicit Ending' as one in which the player performs a specific and explicit gesture with the intention of ending the tone, such as by damping the resonator, by releasing the keys of a piano, or by stopping the bowing or blowing that the musician was engaged in. This gesture requires some cognitive planning, and by our definition, the 'Ending' period begins when the player starts to formulate a motor-action plan about how and when to end the tone. The 'Ending' period reaches its conclusion when the instrument is no longer making

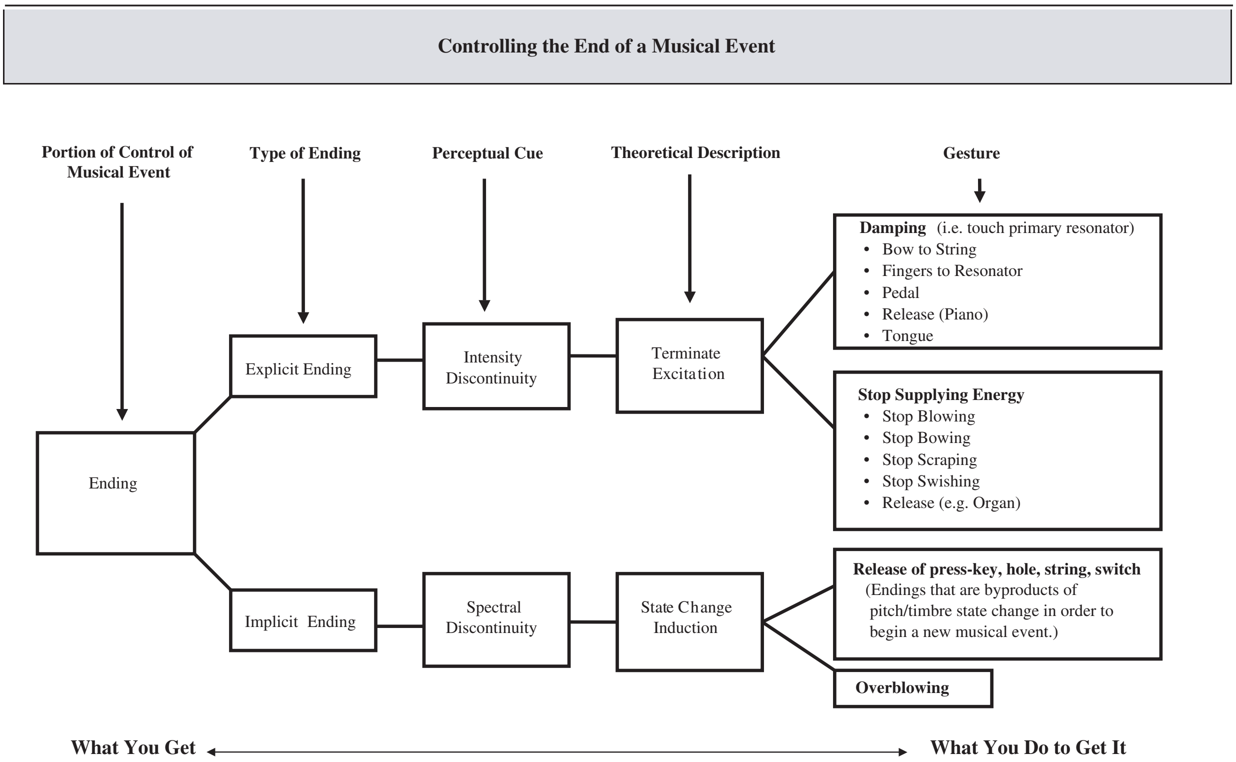


Figure 5

any sound. The perceptual cue for an Explicit Ending is an intensity discontinuity (mirroring the intensity discontinuity that defined an Explicit Beginning). The way in which this is manifested physically is that the source of excitation to the resonator must be terminated. There are two classes of gestures that accomplish this. First are the various forms of 'damping', which we define as touching the primary resonator in some fashion, such as placing the bow on the string, putting fingers or another body part on the resonator, pressing or releasing a pedal. The second class of gestures are the various ways we stop supplying energy to the instrument, either because we stop blowing, stop bowing, stop scraping (as in a gourd or guiro), stop swishing (as in a container filled with black-eyed peas), or by changing the state of a switch/releasing a key (as in the electric organ).

Recall that earlier we made a formal distinction between the *process* of ending a tone and the instant in time when that process is actually *completed*. The moment at which the instrument is no longer producing sound signifies that this process has been completed, and (as described earlier) we call this the *Terminus*. (Since the Terminus is technically not part of the Ending, it is not shown in figure 5.)

For the case in which the player is going to implicitly begin a new tone, we define the 'Implicit Ending' as that portion of a tone when the player is planning how she is going to make the transition from this tone to the next. Technically speaking, an Implicit Ending does not comprise any important physical gestures, only cognitive operations, specifically the intent to change tones. We could argue that at a micro level, there are small preparatory moves made by the player, prior to actually changing the tone (as in the case of 'getting ready to lift a finger off of a flute key') and so properly speaking, these micro-gestures are part of the Implicit Ending process, and are the result of the same sorts of motor action plans that prepare any other musical gesture.

Referring back to figure 2, we see that although a Non-Excited Middle can sometimes go to an Ending (as when we damp a guitar string), it can also bypass the Ending and go directly to a Terminus. In fact the most natural way to end a 'Non-Excited Middle' is to simply do nothing and let it die down by itself. This is, of course, the customary way that drums, cymbals and bells are played, and that is what is meant by the NEM following a logical path directly to the Terminus in figure 2, and (as we mentioned before) the downward slope of the box around the NEM is meant to convey a gradual decay in energy. From a gestural standpoint, the player does nothing and the tone eventually ends. There is a parallel for this gestural description in the perceptual domain. At some point the energy has decayed sufficiently so that the resonator is no longer vibrating (perceptibly) and we hear that the Terminus has been reached.

7. COMBINING THE ELEMENTS OF THE MODEL

How do all these elements come together? In Figure 2, we presented a relatively bare-bones flow chart showing the five highest-level stages of a musical event: the Selective Cognitive Preconditions, the Beginning, Middle, Ending, and the Terminus. In figures 3–5, we elaborated the various ways in which Beginnings, Middles and Endings can be controlled. Specifically, we made a distinction between 'Explicit' and 'Implicit' Beginnings, and within the category of Explicit Beginnings we made the distinction between 'Continuous Excitation Beginnings' and 'Impulsive Excitation Beginnings'. We made a distinction between 'Continuous Excitation Middles' and 'Non-Excited Middles' and finally, a distinction between 'Explicit Endings' and 'Implicit Endings'.

So far, our three charts that deal with the individual parts of a tone body (figures 3–5) are just isolated analyses. It would be useful if there were some principled way in which we could connect them together, that is, if it turned out that 'impulsive excitation beginnings', for example, led only to a certain type of middle; or if 'Non-Excited Middles' could only lead to certain types of ends. It turns out that this is indeed the case. Figure 6 represents a more fully elaborated version of figure 2, one that maps out all the possible paths for a sound, and links together figures 3–5.

Thus, we find that 'continuous excitation beginnings' (CEBs) can lead both to CEMs and NEMs. An 'impulsive excitation beginning' (IEB) can *only* lead to a NEM by definition, because continuous excitation cannot be applied for the first time during a middle. Examples of gestures and traditional musical instruments that illustrate these diagrams are:

- CEB to CEM: bowed violin, flute, scraped gourd,
- CEB to NEM: violin bowed in a martelé fashion, and
- IEB to NEM: drum, cymbal, bell.

Furthermore, we find (in figure 6) that both types of middles can go to both types of endings. For example:

- *CEM to Explicit Ending*. Stop bowing a violin; stop blowing a flute,
- *CEM to Implicit Ending*. Continue bowing a violin and change finger position; continue blowing a flute and change fingerings,
- *NEM to Explicit Ending*. Damp a drum; damp a cymbal, and
- *NEM to Implicit Ending*. Hammer on or pull off a tone on the guitar.

A complete list of all the possible paths is shown in the table.

In figure 7 we expand figure 3 to indicate which instruments (according to the traditional classification system) are controlled by which gestures during the beginning of a tone (c.f. Sachs and Hornbostel 1914/1961, Vertegaal and Eaglestone 1996). This is an elaboration of the right-most column in figure 3.

8. CONTROL AND MAPPING

Given the framework described herein, we can now talk about building new controllers, and about the mappings between gestures and the control of musical sound. By 'mapping' we mean the linking or correspondence between gestures or control parameters (such as we have outlined here) and sound generation or synthesis parameters (Kantowitz and Sorkin 1983, Cadoz, Luciani and Florens 1984, Winkler 1995, Paradiso 1997, Hunt, Wanderley and Kirk 2000, Wanderley 2001b). Mappings should be intuitive insofar as possible (Norman 1988). That is, they should exploit some intrinsic property of the musician's *cognitive map* so that a gesture or movement in the physical domain is tightly coupled – in a non-arbitrary way – with the intention of the musician.

Consider as an example a balance control for a music system. An intuitive gestural mapping would employ a linear fader moving on the horizontal axis: moving the fader control to the left would increase sound on the left, moving it to the right would increase sound on the right. Many non-intuitive balance controls exist, such as linear faders *moving vertically* (which is left: up or down?) or rotary potentiometers (while the top of the rotary knob moves to the right, the bottom of the knob moves to the left – the choice between attending to the top or bottom of the knob's motion is arbitrary; such a mapping is not intuitive and has to be learned – see Levitin, Robson, Smith and O'Sullivan 1996, Norman 1988).

Our existing cognitive maps for sound production include such notions as 'harder means louder' (for breathing or striking), 'gestural wiggling yields pitch or timbre wiggling' (such as in creating vibrato on a stringed instrument), and 'tighter means higher in pitch' (such as when stretching a membrane on a drum, or tightening the embouchure on a wind instrument). These are not arbitrary conventions, but the product of our having evolved brains that incorporate certain specific physical principles of the world around us (Shepard 1994, 1995). Intuitive mappings exploit such principles, and the music produced by them retains a perceptual index specifically related to the gesture (Cadoz 1988). However, not all intuitive gestural mappings have yet been incorporated into traditional acoustic instruments, leaving many opportunities for innovation in the design of computer music controllers.

As one example, the Lightening (Rich 1991, Buchla 2002) allows for an especially wide variety of gestures. If a user rotates his hands in a circle, what should that do in terms of sound control? The circle in and of itself suggests a loop, perhaps a sequence of tones or timbral changes that repeat. An essential idea based on our taxonomy concerns whether the musician gestures a single circle or several circles. What might this gesture mean at the *Beginning* of a musical tone? The single circle suggests a single event: an *Impulsive Excitation Beginning*. Two or more circles suggest, gesturally, that the sound should continue while the hands are in motion: a *Continuous Excitation Beginning*. During the *Middle* of a tone (a *Continuous Excitation Middle*), one can imagine that drawing a circle in the

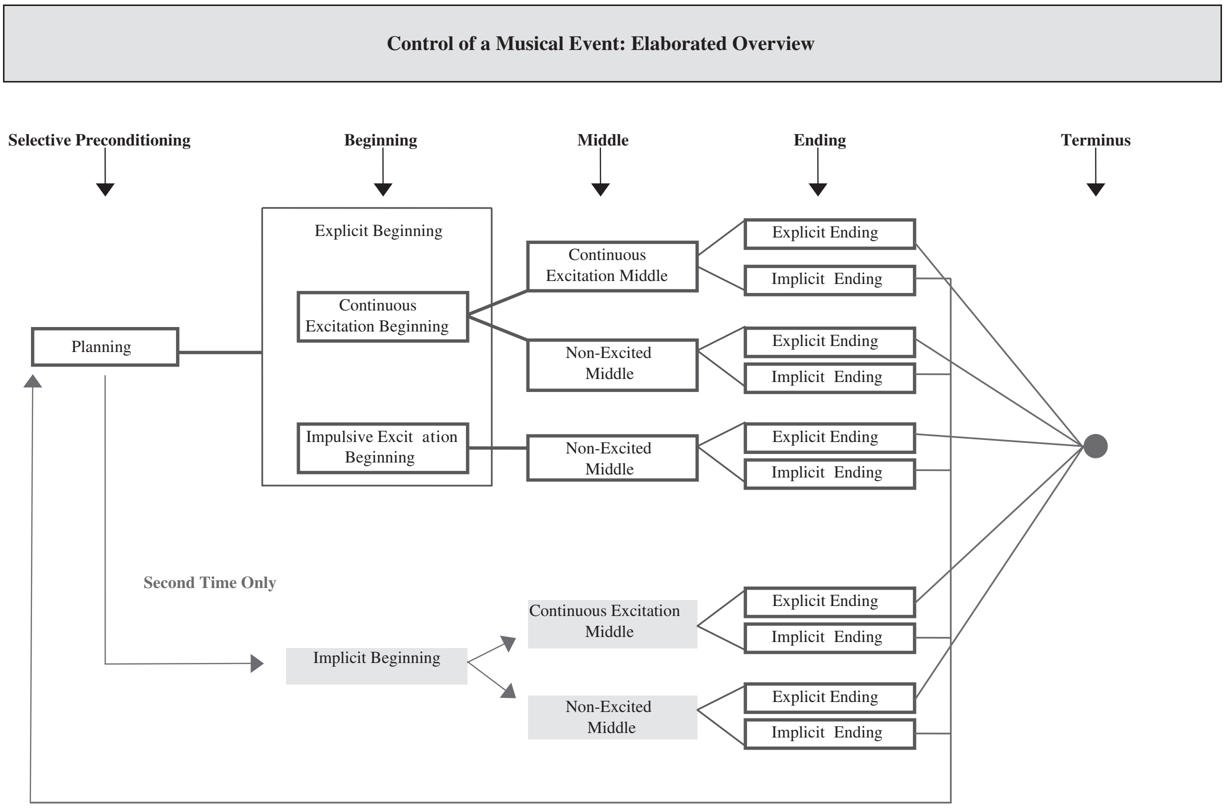


Figure 6

Inventory of gestural ways to create the beginning of a tone

Type of Beginning	Type of Energy	Gesture	Woodwinds	Brass	Strings	Percussion	Keyboards
Explicit	Continuous	Bowing			✓		
		Rubbing				✓	
		Scraping			✓	✓	
		Swishing					
		Blowing	✓	✓			
		Shaking				✓	
Im	pulsive	Hit			✓	✓	✓
		Pluck			✓	✓	
		Press	✓	✓	✓	✓	✓
Implicit	State Change Induction	Release	✓	✓	✓		
		Lift	✓	✓	✓		

Figure 7

air might indicate a circular/periodic manipulation of pitch, loudness and/or timbre.

Most real acoustic instruments can be plucked, struck, tapped, rubbed, bowed, muted, and so on with one or both hands. Musical controllers typically have two parts – the sensors we touch and the mappings between them and the sound module. One thing we can do here with respect to the sensors is to try to emulate the real instruments. For example, press and release a key, as on a piano or wind instrument, bend a string as on the guitar. We might also want to incorporate expressive gestures from other domains (e.g. dance) or draw inspiration from the ways in which non-human animals create sound (Smyth and Smith 2001, 2002).

But what can we learn about *mappings* from the traditional instruments? To make an expressive instrument, it is important to provide options, but not too many. A successful controller will enable the user to follow a well-understood path – in our figure 6, we want to follow a particular path through the anatomy of a tone – an instrument that allowed one to follow *all* possible paths would be too difficult to play.

Although electronic synthesizers offer an increasingly wide range of interesting sounds, instrument controllers

Table. List of all ten logical paths for tone control (derived from figure 8). Key to abbreviations: CEB = Continuous Excitation Beginning, CEM = Continuous Excitation Middle, EB = Explicit Beginning, EE = Explicit Ending, IE = Implicit Ending, NEM = Non-Excited Middle.

EB (CEB)→CEM→EE→Terminus
EB (CEB)→CEM→IE→Beginning
EB (CEB)→NEM→EE→Terminus
EB (CEB)→NEM→IE→Beginning
EB (IE)→NEM→EE→Terminus
EB (IE)→NEM→IE→Beginning
IB→CEM→EE→Terminus
IB→CEM→IE→Beginning
IB→NEM→EE→Terminus
IB→NEM→IE→Beginning

do not typically afford the same range of gestures as acoustic or electrified acoustic instruments (Cook 2002). As we strive to expand the range of gestures we use to control electronic sounds, it is important to avoid the trap of generating sounds that are too uniform – the rigid mechanical sounds so often associated with computers and sequencers. As Risset and Mathews (1969) pointed out over thirty years ago, real musical instruments are not typically operated by robots, 'but by human musicians who introduce intricacies, both intentional and unintentional'. Even if a player wanted to, it is unlikely that (s)he could produce two tones identically, and it is usually not desirable to do so. One of the most challenging goals therefore in the design of computer music controllers is to a climate of individual expressivity and musical nuances, rather than quantising and otherwise impoverishing these gestures with the rigidity and strict repeatability that the digital computer so easily produces (Levitin and Adams 1998).

8.1. Example implementation: The Stick, MAX and the Yamaha VL1

One way in which our research group has implemented some of these ideas has been in the development of an electronic musical instrument we call 'The Stick' (Adams *et al.* 1996, Adams *et al.* 1999, Smith *et al.* 2000), currently played by the musicians Michael Brook and Laurie Anderson (figure 9). The primary objects of The Stick are: (i) to provide an electronic instrument which allows the user to create and/or control musical sounds and other acoustic effects with a high degree sonic realism, (ii) to allow the user to simultaneously and continuously modify many of the parameters of the created sound, (iii) to provide the user with a realistic playing experience, and (iv) to provide an electrical instrument which is easy to master without extensive training and practice, providing inexperienced users with the pleasure of creating music, and experienced users with a high degree of satisfaction and nuanced control.

Although many combinations of hardware devices are possible, in our preferred configuration, The Stick uses digital physical modelling from the Yamaha VL1-M™ Virtual Tone Generator and control signal mapping from MAX MSP (Puckette 1991, Cycling74 2002), a combination we have found to be quite rich in the degree of flexibility it offers. (In fact, the ideas proposed in this paper were inspired by attempts to control physical modelling-based synthesis). Control signals are generated by three collinearly arranged force-sensitive resistors (FSRs), parsed and processed by MAX, and they then access predefined parameters in the VL1. The synthesis module in this arrangement is capable of receiving continuously changing control parameters in real time. The user may select which control parameters of the VL1 are controlled by each of three sensor elements (with pressure and location being sent as MIDI signals).

Our preferred implementation derives from the control and cognitive limitations we've previously mentioned in this paper, and allows the player to have simultaneous control over several parameters of the sound with a single gesture (one-to-many mapping). This is more natural and is consistent with the way real musical instruments behave, and is consistent with psychological theory: because timbre space is multidimensional (Grey 1975, 1977) and involves the perception of correlated dimensions, it makes sense to attempt to control two or more dimensions with one motion – this is, after all, what we do with *real* musical instruments (c.f. Wessel 1979, Lee and Wessel 1992). For example, in an acoustic wind instrument, changes in loudness are accompanied by changes in timbre automatically, as the player blows harder. The Stick (in our preferred implementation) responds in like manner, accomplishing

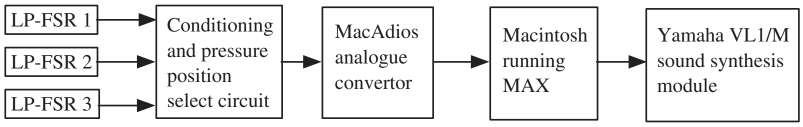
How is pitch selected/created during the beginning of the tone?

Parameter of the energy source being controlled	Physics/type of instrument	Gesture
Length	Vibrating column of air	Press to close or open holes
		Press to close or open or lengthen tubes
	Vibrating String or bar	Press or hit to select desired length
Position	Electronic Keyboard or electronic controller	Press
Tension	Membranophone - vibrating membrane	
	Wind instruments	change uscle/lip tension
	Stringed instruments	change string tension (e.g. by bending)

Figure 8

9a. Michael Brook demonstrates the playing position for The Stick.

9b. Closeup of linear potentiometer FSR on The Stick.



9c. Functional diagram of The Stick

Figure 9

these same real-world transformations through a single gesture. The VL1 provides access to a large number of synthesis parameters such as *pressure, embouchure, tonguing, breath noise, scream, throat formant, dampening, absorption, harmonic enhancer, dynamic filter, amplitude, portamento, growl* and *pitch* (Adams *et al.* 1999, Cook 2002). The one-to-many mapping scheme allows six gestures (pressure and position on each of the three FSRs built into The Stick) to control these parameters in repeatable and expressive ways. In one configuration, the top FSR affects changes in timbre, and squeezing the FSR allows for navigation through a predefined timbre space (squeezing harder produces a more pinched, 'tight' timbre, for example). *Position* on a second FSR controls *pitch*, with *squeezing* on that same FSR controlling *vibrato* (more squeezing yields either faster vibrato or greater vibrato depth, depending on the particular mapping used). From among many possibilities, the key is to choose those that are intuitive and satisfying to the performer. The Stick can be used to control any aspect of sound production through custom mappings, but one way in which we found it to perform especially well is when it allows the musician to navigate through timbre space – that is, when the user can use the FSR sensors to create timbral changes in an ongoing sound.

Our intent with this article has not been to prescribe how specific gestures might map to the control of specific types of sound or sound manipulations, and so we will not go into detail about such mappings beyond the brief examples just given. Rather, our goal has been to create a new structure for *classifying the creation of musical tones based on how they can be controlled*. We leave it to the designer to think of appropriate gesture-to-sound control mappings based on existing principles of interface design (Buxton 1987, 1995), mapping (Wanderley and Orio, submitted), and sound input controllers (Winkler 1995). The essential points are to employ controls that are both consistent with how the sound might have been created in a traditional musical instrument (so as to exploit principles of cognition and the mind), and to inspire designers to think about the control of sound in ways that go beyond those employed in traditional musical instruments (which are constrained by physical, but not always cognitive factors).

9. CONCLUDING SUMMARY

In this paper, we have tried to make explicit the distinction between *musical controllers* and *musical instruments*. Musical controllers are integrated into most traditional acoustic instruments, but the electronic and computer age has seen the separation of the sound controller from the sound-producing device (Mathews 1963). We believe that this separation affords an opportunity to think about how controllers can be designed, no longer subject to the constraints imposed by centuries of integrated design. As a first step to thinking about how new controllers might be designed, we have presented an analysis of the various ways in which sound is controlled, combining a study of *gesture, resonant properties of the instrument* and *perceptual cues*. We noted certain constraints, in particular those of *cognitive* *load* and *motor processing limitations* that can guide designers in developing new controllers.

Finally, through our diagrammatic interpretation of musical event control (figures 1 – 8), we have described a new scheme for characterising musical control space. We first introduced a way of segmenting a single musical tone into five components that are conceptually distinct from a control standpoint (*Selective Preconditions, Beginning, Middle, Ending, Terminus*). We further elaborated those five components by introducing the idea of explicit and Implicit Beginnings, continuous excitation and Non-Excited Middles, and explicit and Implicit Endings. We believe that this parsing of a single musical tone into separate components has a valid and empirical basis, and that each component makes separate demands on the designer of new instrument control structures.

We have tried to characterise musical control space in a new way, and in so doing, our hope is that we have provided musicians, engineers and designers with a new conceptual foundation for thinking about the control of musical sound.

ACKNOWLEDGEMENTS

The preparation of this work was supported by grants to the first author from the Natural Sciences and Engineering Research Council of Canada (NSERC), Fonds Quebec du Recherche sur la Nature et les Technologies (FQRNT), the Canadian Foundation for Innovation (CFI), and Valorisation Recherche Québec (VRQ), and by Interval Research Corporation. We are grateful to Michael Brook, Perry R. Cook, Mark Goldstein, John R. Pierce, Caroline A. Traube, Bill Verplank and Marcelo Wanderley for helpful comments and discussions.

REFERENCES

Adams, R. L., Brook, M. B., Eichenseer, J. W., Goldstein, M. H., and Smith, G. M. 1999. *Electronic Musical Instrument*. United States of America: Interval Research Corp, Patent number 6,005,181.

Adams, R. L., Brook, M. B., Eichenseer, J. W., and Smith, G. M. (Levitin, D. J., ed.). 1996. *The Stick.* Technical Report, IRC 1996–033. Palo Alto, CA: Interval Research Corporation.

Bonnet, M., Guiard, Y., Requin, J., and Semjen, A. 1994. Mécanismes généraux de la motricité. In M. Richelle (ed.) *Traité de Psychologie Expérimentale.* Presses Universitaires de France.

Boulanger, R., and Mathews, M. V. 1997. The 1997 Mathews Radio-Baton & improvisation modes. In *Proc. of the 1997 Int. Computer Music Conf.*, Thessaloniki, Greece. San Francisco: International Computer Music Association.

Bregman, A. S. 1990. *Auditory Scene Analysis.* Cambridge, MA: MIT Press.

Bregman, A. S., and Rousseau, L. 1991. Auditory intensity changes can cue perception of transformation, accompaniment or replacement. *Bulletin of the Psychonomic Society* **29**: 476(A).

Buchla, D. 2002. www.buchla.com

Buxton, W. A. S. 1987. There's more to interaction than meets the eye: Some issues in manual input. In R. M. Baecker and W. A. S. Buxton (eds.) *Readings in HCI: A Multidisciplinary Approach.* San Mateo, CA: Morgan Kaufman Publishers.

Buxton, W. A. S. 1995. Touch, gesture & marking. In R. M. Baecker, J. Grudin, W. A. S. Buxton and S. Greenberg (eds.) *Readings in Human Computer Interaction: Toward the Year 2000.* San Francisco: Morgan Kaufmann Publishers.

Cadoz, C. 1988. Instrumental gesture and musical composition. In *Proc. of the 1988 Int. Computer Music Conf.*, Den Haag, Netherlands. San Francisco: International Computer Music Association.

Cadoz, C., Luciani, A., and Florens, J. 1984. Responsive input devices and sound synthesis by simulation of instrumental mechanisms: The Cordis system. *Computer Music Journal* **8**(3): 60–73.

Chamberlin, H. A., Jr. 1976. Experimental Fourier series universal tone generator. *Journal of the Audio Engineering Society* **24**(4): 271–6.

Chowning, J. 1973. The synthesis of complex audio spectra by means of frequency modulation. *Journal of the Audio Engineering Society* **21**: 526–34.

Cook, P. R. 1999. *Music, Cognition, and Computerized Sound.* Cambridge, MA: MIT Press.

Cook, P. R. 2001. Principles for designing computer music controllers. Paper presented at the NIME Workshop, Chi, Seattle, WA, April 1–2.

Cook, P. R. 2002. derStick: A controller by geeks, for geeks. Unpublished manuscript.

Cycling74. 2002. MAX-MSP. http://www.cycling74.com/products/dlmaxmsp.html

Dawson, M. R. 1991. The how and why of what went where in apparent motion: Modeling solutions to the motion correspondence problem. *Psychological Review* **98**(4): 569–603.

Dennett, D. C. 1991. *Consciousness Explained.* Boston: Little, Brown and Company.

DiGiugno, D. 1976. A 256 digital oscillator bank. In *Proc. of the 1976 Int. Computer Music Conf.*, Cambridge, MA. San Francisco: International Computer Music Association.

Fitts, P. M. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology* **47**: 381–91.

Fitts, P. M., and Posner, M. I. 1967. *Human Performance.* Belmont, CA: Brooks/Cole.

Grey, J. M. 1975. *An Exploration of Musical Timbre using Computer-based Techniques for Analysis, Synthesis and Perceptual Scaling.* Unpublished Ph.D. thesis, Center for Computer Research in Music and Acoustics, Stanford University.

Grey, J. M. 1977. Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America,* **61**(5): 1,270–7.

Hiller, L., and Ruiz, P. 1971a. Synthesizing musical sounds by solving the wave equation for vibrating objects: Part 1. *Journal of the Audio Engineering Society* **19**(6): 462–70.

Hiller, L., and Ruiz, P. 1971b. Synthesizing musical sounds by solving the wave equation for vibrating objects: Part 2. *Journal of the Audio Engineering Society* **19**(7): 542–50.

Hunt, A., Wanderley, M. M., and Kirk, R. 2000. Towards a model for instrumental mapping in expert musical interaction. Paper presented at the International Computer Music Conference, Berlin.

Kantowitz, B. H., and Sorkin, R. D. 1983. *Human Factors: Understanding People–System Relationships*. New York: John Wiley & Sons.

Keele, S. W. 1973. *Attention and Human Performance*. Pacific Palisades: Goodyear Publishing.

Kolb, B., and Whishaw, I. Q. 1990. *Fundamentals of Human Neuropsychology* (3rd edn). New York: Freeman.

Kurtenback, G., and Hulteen, E. A. 1990. Gestures. In *The Art of Human–Computer Interface Design*. Reading, MA: Addison-Wesley.

Lee, M., and Wessel, D. 1992. Connectionist models for real-time control of synthesis and compositional algorithms. *Proc. of the 1991 Int. Computer Music Conf.*, pp. 277–80, Montreal, QC, Canada. San Francisco: International Computer Music Association.

Levitin, D. J., and Adams, R. L. 1998. *Computer-Interaction Design and Controllers: Creating Meaningful Experience*. Technical Report, IRC 1998–005. Palo Alto: Interval Research Corporation.

Levitin, D. J., Robson, D., Smith, G., and O'Sullivan, D. 1996. *El Podiomo – The Multi-Object Podium: A Search for Better Handles for Digital Audio*. Technical report, IRC 1996–025. Palo Alto: Interval Research Corporation.

Markison, R. E. 1990. Treatment of musical hands: Redesign of the interface. *Hand Clinics* **6**(3): 525–44.

Marrin-Nakra, T. 2000. *Inside the Conductor's Jacket: Analysis, Interpretation and Musical Synthesis of Expressive Gestures*. Unpublished Ph.D. thesis, MIT, Cambridge, MA.

Mathews, M. V. 1963. The digital computer as a musical instrument. *Science* **142**(11): 553–7.

Mathews, M. V. 1991. The radio baton and conductor program, or: Pitch, the most important and least expressive part of music. *Computer Music Journal* **15**(4): 37–46.

Mathews, M. V., and Pierce, J. R. 1989. *Current Directions in Computer Music Research*. Cambridge, MA: MIT Press.

McIntyre, M. E., Schuymacher, R. T., and Woodhouse, J. 1983. On the oscillations of musical instruments. *Journal of the Acoustical Society of America* **74**(5): 1,325–45.

MIDI. 1984. *MIDI Specifications 1.0*. http://www.midi.org

Miller, G. A. 1956. The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review* **63**: 81–97.

Mulder, A. G. E., Fels, S. S., and Mase, K. 1997. Mapping virtual object manipulation to sound variation. In T. Rai and R. Basset (eds.) *IPSJ SIG Notes* **97**(122): 63–8.

Norman, D. A. 1988. *The Design of Everyday Things*. New York: Doubleday.

Palmer, S. E. 2000. *Vision: From Photons to Phenomenology*. Cambridge, MA: MIT Press.

Paradiso, J. 1997. Electronic music interfaces: new ways to play. *IEEE Spectrum* **34**(12): 18–30.

Pierce, J. R. 1983. *The Science of Musical Sound*. New York: Scientific American Books/Freeman.

Puckette, M. 1991. Combining event and signal processing in the MAX graphical programming environment. *Computer Music Journal* **15**(3): 68–77.

Rich, R. 1991. Buchla lightening MIDI controller. *Electronic Musician* **7**(10): 102–8.

Risset, J.-C., and Mathews, M. V. 1969. Analysis of musical-instrument tones. *Physics Today* **29**(2): 23–9.

Roads, C. 1996. *The Computer Music Tutorial*. Cambridge, MA: MIT Press.

Roads, C., and Strawn, J. 1985. *Foundations of Computer Music*. Cambridge, MA: MIT Press.

Sachs, C., and Hornbostel, E. M. 1914/1961. Systematik der Musikinstrumente/Classification of musical instruments. *Galpin Society Journal* **14**: 3–29.

Saldanha, E. L., and Corso, J. F. 1964. Timbre cues and the identification of musical instruments. *Journal of the Acoustical Society of America* **36**: 2,021–126.

Schaeffer, P. 1966. *Traité des objets musicaux*. Paris: Editions du Seuil.

Shepard, R. N. 1994. Perceptual-cognitive universals as reflections of the world. *Psychonomic Bulletin & Review* **1**(1): 2–28.

Shepard, R. N. 1995. Mental Universals: Toward a twenty-first century science of mind. In R. L. Solso and D. W. Massaro (eds.) *The Science of the Mind: 2001 and Beyond*, pp. 50–61. New York: Oxford University Press.

Smith, G. M., Goldstein, M. H., Eichenseer, J. W., Brook, M. B., and Adams, R. L. 2000. *System and Method for Controlling a Music Synthesizer*. United States of America: Interval Research Corp, Patent number 6,018,118.

Smith, J. O. III. 1992. Physical modeling using digital waveguides. Computer *Music Journal* **16**(4): 74–91.

Smyth, T., and Smith, J. O. III. 2001. Applications of bioacoustics in physical nmodeling and the creation of new musical instruments. In *Proc. of ISMA 2001, Int. Symposium on Musical Acoustics*, Perugia, Italy.

Smyth, T., and Smith, J. O. III. 2002. Creating sustained tones with the cicada's rapid sequential buckling mechanism. In *Proc. of the 2002 Conf. on New Instruments for Musical Expression (NIME)*, Dublin, Ireland.

Stumpf, C. 1926. *Die Sprachlaute, Experimentell-phonetische Untersuchungen*, p. 374. Berlin: Springer.

Taylor, C. A. 1992. *Exploring Music: the Science and Technology of Tones and Tunes*. Bristol, UK; Philadelphia, PA: Institute of Physics Publishing.

Verplank, W., Mathews, M.V., and Shaw, R. 2000. Scanned Synthesis. In *Proc. of the 2000 Int. Computer Music Conf.*, Berlin, Germany. San Francisco: International Computer Music Association.

Vertegaal, R., and Eaglestone, B. 1996. Comparison of input devices in an ISEE direct timbre manipulation task. *Interacting with Computers* **8**(1): 13–30.

Wanderley, M. M. 2001a, September 2001. Gestural control of music. Paper presented at the *Int. Workshop on Human Supervision and Control in Engineering and Music, Kassel*, Germany, September.

Wanderley, M. M. 2001b. *Performer–Instrument Interaction: Applications to Gestural control of Sound Synthesis*. Unpublished Ph.D. thesis, Université Pierre et Marie Curie, Paris VI, Paris.

Wanderley, M. M., and Orio, N. 2002. Evaluation of input devices for musical expression: Borrowing tools from HCI. *Computer Music Journal* **26**(3): 62–76.

Wertheimer, M. 1923/1938. Laws of organization in perceptual forms. In W. D. Ellis (ed.) *A Source Book of Gestalt Psychology*. London: Rutledge & Kegan Paul.

Wessel, D. 1979. Timbre space as a musical control structure. *Computer Music Journal* **3**(2): 32–45.

Winkler, T. 1995. Making motion musical: Gesture mapping strategies for interactive computer music. Paper presented at the *Int. Computer Music Conf.*, Banff, Canada.