# Digital Music Encoding as Cultural Practice

Richard J. Lewis

Thesis submitted to the University of East Anglia for the degree of Master of Music

December 2008

# Abstract

**Digital Music Encoding as Cultural Practice**

**Richard Lewis**

**2008**

We examine the use of computers in storing and manipulating music. We consider the validity of treating music as information in the formal terms required by computers. We take the metaphor of inscriptions (marks on a medium) and draw out its implications for music representation techniques and digital encoding practices through its relationship to notations and to digital storage, its ability to take on semantics and become a representation, its ability to be gathered together into documents, and its ability to be disseminated, particularly over digital networks. We then examine some examples of practice in designing and applying digital music encoding methods and draw some conclusions for the practice of computer assisted musicology: that suitable encoding methods are vital for any application of computers in music research, and that users must understand how musical information is being represented in order to make optimal use of the techniques.

# Contents

## Acknowledgements

My thanks are due to Jonathan Impett and Simon Waters who both contributed helpful and motivating discussion throughout my research. Also to Zoë Rowsell for proof reading my text. Any remaining mistakes and inaccuracies are my own.

# Introduction

There is a thriving culture of scholars from musicology and computer science disciplines who devote their attentions to examining the phenomena of sound and music (in many of their manifestations) with the aid of information technology. What all such practices have in common is that they rely on being able to treat music as information and being able to abstract it into a form suitable for digital[1] storage and manipulation. Cook (2005, 1) describes this as the "interdisciplinary interface" of "music information science". A mixture of high productivity and easy dissemination (through digital networks) has led to a proliferation of such practices.

Huron describes how, "[i]n the initial decades of humanities computing there was a flurry of activity focussing on the problem of representation." (Huron, 1992, 5) He goes on to propose some "lessons" on the issues surrounding designing music representation methods: "1. The types of goals to which music representations may serve are legion and unpredictable. 2. Information which is crucial to the performance of one task is likely to be irrelevant in the performance of other tasks. 3. One must resist the tendency to assume that the issue of computer-based music representation is primarily one of representing notated scores. 4. Our ignorance of how users intend to utilise a representation (either in the immediate or distant future) limits our ability to design an optimum scheme." While his first three are clearly important in the handling of music information, it is his last lesson which touches upon the importance of music encoding as a cultural practice.

It will be our hypothesis that these encoding methods succeed (or fail) to the extent that they manage to fit into a cultural context. This work, then, seeks to explore the validity of such a hypothesis by examining what is involved in encoding musical information digitally and surveying some existing practices, both strands of investigation being tempered by the question of how far the subject acknowledges and seeks to situate itself within a cultural context.

As we shall see, such cultural contexts include practices such as creative use of sound, music analysis, archiving and cataloguing, and formalised approaches to music theory, many of which may benefit from mixing the disciplines of computer science and music. Camilleri (1992, 171–172) argues that interdisciplinarity, particularly in methodology (and

[1] We use "digital" as a convenient (and reasonably fashionable) term to demarcate any practices which deal with information stored in (principally) electronic form and manipulated and communicated with the aid of computers.

even more specifically in computational approaches to "music theory"), can lead not only to adoption of new "terminology, strategies and concepts", but also to reformulations of core theoretical principles.

We will examine the nature of inscription and what it means in terms of sound and music and how it is manifest digitally. We will examine the process of representation, how one thing can be seen as a sign for another, and how this property is utilised in music and in computers. We will examine the musical work as an intellectual entity and its manifestation in the form of a document (as well as saying about the *document* itself). Finally, we will examine by categorisation a few examples of practice in the field of making digital music encodings.

We begin by considering what Bruno Latour has to say about the nature and use of inscriptions. His ideas will become central to the whole discussion and focus on the idea of *displacement*, how one thing (e.g. an inscription) displaces another (e.g. a real-world object) by being a representation for it and pushing it away as an unnecessary complexity. He argues, "[w]ithout the displacement, the inscription is worthless; without the inscription the displacement is wasted." (Latour, 1990, 16) Every attempt to address the music requires a displacement. The music cannot be pinned down, cannot be held in the hand, it is always displaced. Following Latour, this inevitable displacement is "wasted" if not inscribed somehow. Taking this as a point of departure, we attempt to understand how scholars have displaced sound and music and inscribed their resulting representations using digitised techniques.

# Chapter 1

# Inscription

## 1.1 Inscriptions

The instigator and probably most celebrated (or at least most famous) exponent of the field of science studies, Bruno Latour, has a remarkable reverence for the seemingly plain and inconspicuous practice of inscription. Of inscriptions as things he says they are, "*mobile* but also *immutable*, *presentable*, *readable* and *combinable* with one another" (Latour, 1990, 6) and argues that they have an extraordinary power to mobilise, to bring numerous times and numerous places into one time and place: the inscribed object. An economic analogy may be found in the concept of money: goods and services rendered need never be moved around, it suffices simply to represent their value numerically and inscribe that value as promises on pieces of paper, or numbers stored in far away computer databases.

As well as the medium of communication, Latour describes the inscription as, "*the fine edge* and *the final stage* of a whole process of mobilization, that modifies the scale of the rhetoric. Without the displacement, the inscription is worthless; without the inscription the displacement is wasted." (Latour, 1990, 15–16) This chapter, then, will deal with some of these properties of inscriptions as they relate to the writing down of music, to storing information in computers, and to encoding musical information digitally. Latour's concept of "displacement" will become central in the following chapter.

#### Cascading

For Latour, the making of inscriptions is never a complete work. Although each individual inscribed instance may be "immutable", it forms part of a "cascade" of inscriptions which are continually re-made, re-inscribed. He describes the management of a museum collection. The collection is too large for any one curator to be able to understand in completeness and so it is continually catalogued and re-catalogued, extending, Latour argues, “the process through which the specimens had been extracted from their contexts.” (Latour, 1990, 16)

A corollary to this cascading process is to be found in musical inscription: musical notation “extracts” music from its performative or originating compositional “context”, and digitised inscriptions (re-inscriptions or transcriptions) “extract” paper (manuscript) inscriptions from their “context”. At each stage of the process information may be lost or gained either accidentally (for example, by omission or misinterpretation) or deliberately (for example, again by omission, or by augmentation or annotation).

Latour takes up this idea of changing informational content when he describes how each successive inscription in such cascades tends to become more simple than its source inscription. He gives examples including meteorological photographs being re-inscribed by computers as simple line drawings, and DNA being re-inscribed as grey-scale images, and further as strings of letters from a four-letter alphabet. In each case, the new inscription loses detail but gains clarity, it becomes a more informative representation.

As well as becoming more simple (or perhaps even as a manifestation of a class of greater simplicity), Latour argues that cascading inscriptions, “always move [in] the direction of the greater merging of figures, numbers and letters, merging greatly facilitated by their homogeneous treatment as binary units in and by computers.” (Latour, 1990, 16) Similarly, on the subject of the digitisation of media, Lunenfeld argues that the digital is becoming “the universal solvent into which all difference of media dissolves into a pulsating stream of bits and bytes” (Lunenfeld, 1999, 7). For Latour as well as for Lunenfeld, digital storage seems to be a medium as blank and unassuming as plain paper, a syntactically neutral substrate on which any semantics may be implemented. However, while Latour celebrates the versatility of the digital medium, Lunenfeld and other media theorists seem almost to fear its homogeneity as if it were a threat to the richness of diverse media. We shall return to this discussion of the relationship of technology and inscription later.

### Properties of Inscriptions

Returning to inscriptions in general, Latour lists nine advantages of using an inscription of a thing over using the thing itself which are all oriented around what he calls "accelerating" and "summarizing" a "mobilization process" (Latour, 1990, 20).

The first is just that: inscriptions are "*mobile*". It's possible to move a map of a terrain, but not the terrain itself. A photograph will capture a moment in time at a particular place and allows that moment to persist over time and to be carried around to other places. Similarly, a musical score or recording mobilises a musical work which would otherwise only be available in performance or recollection.[1]

Inscriptions are "*immutable*", once made they don't change. Far from precluding the re-inscription process Latour describes as cascading, the immutability of inscriptions actually aids it in that an immutable inscription is always available as a source for re-inscription.

Inscriptions are "*flat*", especially on a paper medium they have no hidden sub-parts, the whole inscription is visible at once. This, Latour argues, is the chief advantage an inscription gives in allowing practitioners and scholars (or scientists in his case) to "master" or "dominate" their field, that is, to gain a full and complete understanding of it.

Inscriptions are *scalable*. Unlike the real-world objects they represent, inscriptions can be re-scaled to different sizes without changing their "internal proportions".

Inscriptions are "*reproducible*". It is much easier to make a copy of an inscription of a thing, than to make a copy of the thing itself. This property greatly facilitates the "mobilization" process which Latour describes. Both printing and digital copying (as we shall see later) are prime examples of how reproducible inscriptions can be.

As a result of being mobile, immutable, flat, scalable, and reproducible, Latour argues that inscriptions are "*recombinable*". He describes what he calls an "optical consistency" which allows inscriptions to be "reshuffled" and, he notes, forms the basis of metaphor.

Furthermore, Latour argues, just as inscriptions may be recombined, they may also be "*superimposed*", layered on top of one another. In this way, inscriptions can flatten the differences between disparate disciplines or phenomena; it was through the inscriptions of cantometrics, the maps and tables of figures, that Alan Lomax was able to link geography with musical style (Lomax, 1976). Latour also describes how, "[m]ost of what we call

[1] Arguably, recollection may involve some form of storage—representation and inscription—upon some kind of substrate.

'structure', 'pattern', 'theory', and 'abstraction' are consequences of these superimpositions." As we shall see later, these concepts are central both to digital paradigms and to contemporary musicology.

Another property of inscriptions crucial to the practice of musicology is, as Latour describes, that they can "after only a little cleaning up" be "*made part of a written text*". He argues that as the text and the inscribed objects take on the same "optical consistency" and "semiotic homogeneity", the objects, and the discourse with which the text engages (through citations and references) actually become part of the text, they are mobilised into it. "The text is not simply 'illustrated', it carries all there is to see in what it writes about. ... the text and the spectacle of the world end up having the same character." (Latour, 1990, 20) In musicological discourse, this concept is borne out in the fact that musical inscriptions (e.g. notation) may be included in and referred to by the text. Notation makes music something you can talk about and thereby at least partly solves the problem summed up in the quote often attributed to the guitarist and song-writer Elvis Costello, "writing about music is like dancing about architecture".[2]

Finally, what for Latour is the "greatest" advantage of inscriptions is that, as a result of their "two-dimensional character", they may be "merge[d] *with geometry.*" It's no longer necessary to manipulate or measure the real-world objects, only their inscriptions need be worked with. Some musical inscription practices take advantage of the two-dimensional character of the flat surface, making it continuous with two chosen dimensions of music, most often (see for example common Western notation) pitch and time. Musicologists (and composers and arrangers) can then manipulate the musical or sonic material *through* the inscription (either on paper or on the computer screen) without dealing with the sound directly at all. In fact, a lot of the more traditional instruction in compositional techniques (harmony and counterpoint) is about how to understand and manipulate musical notation as a surrogate for sound; notice how harmony examinations are conducted in silence.

Latour is not alone in celebrating these properties of inscriptions. Even Nelson Goodman, in his account of the function of notations, admits of what he calls the score's "exciting" utility in "facilitating transposition, comprehension, or even composition" (although, as we shall see later, he describes this utility as subordinate to "its logically prior office" of "identifying a work" (Goodman, 1976, 128)). Similarly, Nicholas Cook describes the score

[2] http://www.pacifier.com/~ascott/they/tamildaa.htm Accessed 9 May 2009

as "highly reduced" and as "data you can actually manipulate" (Cook, 2005, 3), arguing that this is something musicologists find very appealing about working with notated scores as opposed to any other representation of music.

## 1.2 Inscribing

Having considered something of the nature of inscriptions as things, we now expand upon some of the points Latour makes about the uses of inscriptions and turn to the process of making them, of inscribing.

#### Inscription and Understanding

One of the founders of the theory of semiotics, Charles Sanders Peirce, describes a taxonomy of signs consisting of "iconic" signs (those which resemble the thing they represent), "indexical" signs (those which stand in a literal or conceptual *pointing-to* relationship with what they represent, e.g. smoke indicating fire), and "symbolic" signs (those in which the relationship between the sign and the thing represented comes about as a result of convention) (Peirce 1931–58, cited by Chandler 2002, 36–37). Chandler (2002, 45) also describes how sign systems tend to develop from the iconic or indexical to the symbolic, they become more conventional and less representative through use. Obviously, this shifting relationship is reflected in how users engage with the sign system—the conventions must be learned and actually restrict and dictate what is expressible within the system. "Language plays a crucial role in 'constructing reality' ", Chandler argues. He describes Ferdinand de Saussure's stance that reality is a "seamless continuum" which is divided into arbitrary concepts through language, "where, for example," he asks, "does a 'corner' end?" (Chandler, 2002, 27) Just as languages cause their users to view the world in a particular way, so do inscriptive systems dictate how the inscribed phenomenon is understood.

So, far from being a simple labelling system, an inscriptive method actually defines a practice. Latour gives the example that chemistry is not a science while its practitioners are only manipulating physical substances. Rather, it "becomes powerful only when a visual vocabulary is invented that replaces the manipulations by calculation of formulas." (Latour, 1990, 13)*Doing* chemistry is about *doing* chemical formulae.

### Inscription and the Creative Process

For many composers, then there is an important conceptual entanglement between the creative process of composing music and the inscriptive process of writing it down. Conventional Western notation (and particularly technologised versions of it—think about copying and pasting in your notation editor), graphic scores, tape music composing techniques, and software packages for working with sound all impose restrictions on how sound can be manipulated which are a product of their inscriptive characteristics, rather than any inherent restrictions in the nature of the sounds manipulated. Similarly for music analysts, choices about interpretation of musical material depend as much on the inscriptive method employed in describing those interpretations (as well as how the music to be analysed is inscribed) as they do upon any sort of pre-inscriptive insight into the nature of the music in question. Schenker graphs provide a means for inscribing prolongations, so Schenkerian analysts find prolongations in the music they study.

The composer Roger Reynolds (2005), in his description of the use and design of notations, establishes three "creative continua" on which any inscriptive practices in musical composition depend. The first is the path of "decision-making" from the "creator's idea to the listener's response": "idea, crystallization, storage, study, performance, hearing, and interpretation." The second is a five-stage continuum "hesitantly borrowed from Information Theory: source, transmitter, channel, receiver, and destination". And the third sees the compositional idea as "*active*", the notation as "*medial*" (a kind of transmission mechanism or medium), and the performance as a "*passive*" (Reynolds, 2005, 129) result. Of these three, Reynolds argues that the first is most relevant to creative processes enabled by and realised through inscription.

He emphasises the importance of "decision-making" and describes how different kinds of creative practice concentrate decision making at different points of his first continuum (Reynolds, 2005, 131). For example, decision making in through-composed works is most concentrated in the "crystallization" phase, whereas in performing from figured bass or playing indeterminate music, it is deferred to the "performance" stage.

When decision making in the creative process is "deferred" (as Reynolds puts it) to a stage such as his "storage", the act is related to what Latour calls "displacement". The decision made is displaced and becomes immutable by being stored.

Describing this creative continuum (particularly the "crystallization", "storage", and "study" phases), Reynolds explains how, given an idea, the composer then attempts to crystallise it, making it "more manageable, more communicable" (or "mobile" in Latour's terms). Then the idea goes through a process of storage, "encoded in the memory, on paper, or on tape", making it "fixed" so that it may be "considered and *re*considered at will" through both "internal" and "external" "rehearsal" (examples of what Latour calls "recombination"). Similarly, Impett describes composers' use of sketches as the "conscious, externalised keeping track of the evolution of [musical] ideas" (Impett, 2000, 3). Inscribing is an essential part of composing.

In designing new notational systems, Reynolds (2005, 137) argues that composers are likely to "misjudge" "in first attempts at notating unprecedented events or systems". The reasons for this include not only things like finding the right symbol or words to represent "some new aural phenomenon", but also the restriction to just two dimensions of inscription. He argues that this often leads to "unsolvable logical dilemmas" as the number of "sound parameters" to be described often exceeds two. We shall return to these representational issues in the next chapter.

Reynolds also describes the importance of "saying enough to stimulate [the performer] without inhibiting" (Reynolds, 2005, 137) and argues that the use of words in contemporary scores is not uncommon as a last "resort", describing it as an "intermediary stage." Mixing words in amongst what are essentially diastemic symbols raises some questions as to the "optical consistency" of the two types of inscription. As we saw above, Latour describes the ability of text and other inscriptions to be mixed in discourse, and the same is clearly true of mixing musical inscriptions into, for example, music analytic discourse. However, when words are mixed into musical inscription, they act as a kind of stumbling block, an optically inconsistent foreign body not following the dimensionality of their inscriptive host. Of course, Reynolds describes words as a last "resort" and "intermediary" in notations and so may expect any use of them eventually to be flattened and made consistent with dimensionally laid out symbols.

In designing notational systems, Reynolds argues that it's important to keep in mind the notion of "limited channel capacity", that is, "[w]e encode and store—*we comprehend*—by no means all of what we see; the less easily a symbol or idea is mapped into the inner fields of our previous experience, the less likely we are to register it accurately." (Reynolds, 2005, 139) Therefore, he argues, the symbols used should embody sufficient redundancy to negate the possibility of misinterpretation as a result of a "momentarily clouded brow".

So Reynolds gives us some insight into how inscribing enables the creative process of composing music and also warns us that the process is not always easy. His basic model of idea, store, refine, re-store is applicable to other creative processes as well such as writing and drawing. We shall revisit some of his ideas later in the context of a discussion of the nature of musical inscriptions.

## 1.3 Musical Inscription

In what follows, we examine two major classes of musical inscription: notation and recording. Before doing so, however, we briefly examine how Latour's advantages of inscriptions may apply to musical inscription.

*Mobilisation*: it is clear that having a score of a musical work is a very effective transmission mechanism from composer to performer (as long as both are literate in the notational system employed). Similarly, recordings allow (admittedly often idealised) musical performances to be disseminated easily, but again depending on adequate provision of rendering technology.

*Immutability*: paper scores provide the same kind of immutability that Latour will admit of for inscriptions in general (as he is most concerned with ink on paper). Recordings, on the other hand, provide a different kind of immutability, in that, unlike a score, there is not even room for alternative renderings of their musical content. They are immutable in the extreme[3].

*Flatness*: many scores arrange two dimensions of musical continua on the two dimensional surface of the page (often pitch and time). It is then possible to *see* the temporal, serial phenomenon of music at a glance, to flatten its time dimension. Latour also talks about flatness as not hiding any parts of the inscribed phenomenon which, by making a single, unchanging rendering of the music, a recording arguably does.

*Modification of scale*: we discuss some important implications of changing scales in the discussion on digital inscriptions below. For now, though, we can note that notational inscriptions of music, as they admit of re-rendering the music itself, are scalable (for example, performances can be given at different tempi), whereas recordings change their

[3]Of course, this discounts practices such as John Oswald's plunderphonics.

character considerably when their temporal dimension is altered.

*Reproducibility*: music notation is reproducible to the extent that Latour's generalised inscriptions are reproducible. Recordings, too, may be (especially in the case of digital recordings), reproduced with the effect of making exact copies of a rendition of the musical content.

*Recombinability*: the issue of recombination and musical material touches on the work of composers. Composers working with both notation and sound (as audio on tape or stored digitally) have taken advantage of the ability to recombine their inscribed material, often with transformation, in order to create coherent musical structures. (Consider sonata form one large effort of recombination of two fragments of musical material, an effort greatly aided by the facility to notate those two fragments and all their recombined forms.)

*Superimposition*: both notation and recordings admit of some classes of superimposition. For example, aligning parts together to produce scores (especially in early music transcription), and editing multi-channel recordings. There are also digitally enabled research efforts in producing multi-model music representations in which inscriptions such as notation and audio are encoded and presented orthogonally, allowing superimposition such as automatic score position indication in playback.

Latour describes how inscriptions can, "after only a little cleaning up" be "*made part of a written text*". This is certainly true of musical notation and is the reason why it's possible to produce music analytic discourse. As we discuss below, music examples may be included directly in the text which describes them. Recordings, on the other hand, lack what Latour calls "optical consistency" with text and so, excepting such practices as multi-modal representations, do not suit inclusion in written texts well.

Finally, Latour argues that inscriptions may be "merged with geometry". Diastemic notation (such as CWMN) is an obvious example of such an advantage as it is inherently dimensional and distances between elements in the inscription (although not on absolute scales), correlate with the distances in the sonic phenomena they represent. Even recordings (either through mechanical or digital means) allow at least a basic level of measurement such as time, pitch, loudness.

We now go on to consider some conceptions of musical notation and recording in more detail.

### 1.3.1 Notation

One of the most widely used manifestations of musical inscription is notation. Hugo Cole describes the importance of notation in Western practice: "if we consider the established pattern of Western musical activity—in which A tells B what he is to do for the benefit of C—it is clear that the whole process depends on the choice of a suitable notation to serve as a link between A and B; one which will both express what needs to be expressed and allow information to flow smoothly between the two." (Cole, 1974, 1) However, as we shall see below, the power of notation may not reside exclusively in its utility as a means of communication (its "mobility" in Latour's terms) for a cultural practice, but also in its power to displace musical ideas for composers as well as for performers.

Before examining some of the features of musical notations, we will first cover some of the properties of notations in general.

#### Goodman's Theory of Notation

Nelson Goodman, in his theoretical text on the nature of symbol systems in the arts, gives a detailed account of what he calls his "Theory of Notation". In it he defines five requirements of a notational system, two syntactic and three semantic: syntactic character indifference, syntactic character finite differentiation, semantic unambiguity, semantic disjointness, and semantic finite differentiation.

Before elaborating on these requirements, however, it's necessary to understand Goodman's stance on inscription. Goodman employs three terms in describing inscriptions: "marks", "inscriptions", and "characters". Marks carry the least meaning, they are simply any imprint on a medium. Inscriptions are a special class of mark: they are marks which conform to a character, instances of that character. And, consequently, a character is a template for an inscription, as Goodman puts it, "classes of utterances or inscriptions or marks" (Goodman, 1976, 130).

Goodman also draws a distinction between "*atomic*" and "*compound*" inscriptions, "[a]n inscription is *atomic* if it contains no other inscription; otherwise it is *compound*." (Goodman, 1976, 141) However, he also argues that, given an existing scheme to analyse, the designation of an inscription as atomic or compound is often a matter of interpretation rather than a clear distinction. He also argues that there are "virtually no feasible" schemes in which every possible combination of inscriptions is also a valid compound inscription. Rather, inscriptions may only be legitimately combined within the rules of the scheme. Finally, he notes that the principles of compoundness also extend to characters, a compound character is one such that its inscriptions are compound, and that all the requirements for notations apply to compound characters just as they do to atomic characters.

Much like the sign systems of semiotics, Goodman's requirements for notations are only applicable when the inscriptions are working in a context. He distinguishes "schemes" and "systems" roughly along the lines of the syntactic/semantic divide. For Goodman, schemes are purely syntactical, whereas systems are schemes with semantics, "[a] symbol system consists of a symbol scheme correlated with a field of reference." (Goodman, 1976, 143) He also draws a distinction between symbol schemes/systems and notational schemes/systems. It's essentially a simple subsumption relationship: all notational schemes/systems are symbol schemes/systems, but not all symbol schemes/systems are notational, Goodman's requirements of a notational scheme/system are more demanding than those of a mere symbol scheme/system.

Having established some of Goodman's terminology, we can now examine his "syntactic requirements" for notations. The first he describes as "character-indifference". The syntactical function of a particular character in the notational scheme must be able to be fulfilled by any inscription of that character. Two *marks* are character-indifferent if both are *inscriptions* and "neither one belongs to any character the other does not." He describes this relationship as "a typical equivalence-relation: reflexive, symmetric, and transitive." (Goodman, 1976, 131) Goodman also describes a corollary requirement for characters which is more related to their inscriptive manifestation: that they must be "*disjoint*" on the carrying medium. So just as character-indifference ensures syntactic unambiguity, disjointness ensures inscriptional unambiguity. However, Goodman does also admit of the importance of context in determining the character of an inscription, determination is not necessarily a "simple function of shape, size, etc." (Goodman, 1976, 138) of the inscription, but often relies on the character classes of the surrounding inscriptions. He gives the example that the alphabetic characters "a" and "d" could easily be inscribed identically, but their *word* context will determine their interpretation.

The next property Goodman describes of notational schemes is that characters must be "*finitely differentiated* or *articulate*" (Goodman, 1976, 135). That is, it must at least be theoretically possible to tell inscriptions of the characters apart. To illustrate, he gives the example of a notational scheme whose characters are made up of straight lines and in which any difference in length constitutes a different character. Because there are infinitely many possible characters in this notation and because any given inscription will be indistinguishable from any other it is not finitely differentiated. He stresses that finite differentiation "neither implies nor is implied by a finite number of characters." (Goodman, 1976, 136)

Having examined how Goodman believes notational schemes to be constructed, we now go on to consider how those schemes become what Goodman calls "systems" and do the work of displacement. Goodman expresses the relationship between a symbol system and its real world referents as a process of "compliance", the objects denoted *comply* with inscriptions in the system. (Goodman also notes the converse of this relationship, that inscriptions in the system *denote* compliant objects.) Just as he established two classes of inscriptions (atomic and compound), he establishes three classes of "semantic classification": "vacant", "prime", and "composite" (Goodman, 1976, 144). As Goodman describes, compliance with an inscription (or, using Latour's terms, displacement by an inscription) is not limited to one-to-one mappings between inscriptions and objects, but also includes correlations between compound inscriptions and relationships among objects. For example, the linear succession of note symbols in common Western music notation is correlated with the temporal succession of sounds.

Elaborating on these semantic classifications: "vacant" inscriptions are simply those which are syntactically valid but which have no compliant in the field of reference (Goodman gives "green horse" as an example).[4] An inscription is "prime" if no component of it (that is, none of its constituent marks) is also an inscription correlating with an object in the system. If, however, an inscription is made up of other inscriptions which *do* denote objects, then Goodman considers that inscription to be "composite". This is distinct from compound characters as the components have to be *semantically* as well as syntactically valid in isolation in order to be classed composite.

Taking these basic components for constructing semantics from a notation, we now examine Goodman's "semantic requirements" of notations. The first is that a notation should be "*unambiguous*", meaning that "the compliance relationship is invariant". For

[4]Interestingly, Goodman also admits of the notion of an object which complies with no inscription in a symbol system and describes such objects as "unlabelled in the system." (Goodman, 1976, 145)

Goodman, any ambiguity in a notation will mean that the "identity of the work will not be preserved in every chain of steps from performance to covering score and from score to compliant performance." (Goodman, 1976, 149) Similarly, his second semantic requirement is that the "compliance-classes must be disjoint" (Goodman, 1976, 148). Whereas unambiguity requires that the relationship between an inscription and a class of compliant objects must stay the same, disjointness requires that that relationship must be between exactly one inscription and exactly one class of compliant objects. Goodman does, however, allow an object to be compliant with more than one inscription, describing the situation as "redundancy" (Goodman, 1976, 151).

Goodman calls his final semantic requirement "*semantic finite differentiation*". This requirement excludes what he calls "semantically dense" (Goodman, 1976, 153) systems, that is, systems where for any two compliant objects, there is always another compliant object between them (such as for two pitches in a glissando).[5]

So how does Goodman's theory fit into a cultural context? Goodman himself recognises that his five requirements of notational systems are incomplete in that they admit of entirely empty notations as well as making "[n]o requirement of a manageably small or even finite set of atomic characters, no requirement of clarity, of legibility, of durability, of maneuverability, of ease of writing or reading, of graphic suggestiveness, of mnemonic efficacy, or of ready duplicability or performability". Many of these requirements echo Latour's advantages of inscriptions and so highlight an important distinction between those advantages and Goodman's "notations": for Goodman, notation is not much more than a hypothetical theory, an "intricate, abstract, and probably trying technical study" (Goodman, 1976, 157) which, in its purest form, can't possibly have any practical external corollary. Latour, on the other hand, derives his ideas about the nature of inscriptions from observation, he describes the real applications of inscriptions as he has witnessed them.

However, the functioning of displacement is evident in Goodman's theory. Initially on a micro scale: the progression from the most concrete "mark" through "inscription" to the most abstract and displaced "character". It is his conception of the function or purpose of a notation, though, that best exemplifies evidence of displacing. Goodman's insistence

[5]Note the similarity between this and differentiation of *characters*. The difference is that *semantic* differentiation refers to the density of compliant real-world objects, whereas character differentiation is a feature of the inscribed notation itself.

that conformant notations should allow lossless conversion from score to performance and back best demonstrates his belief that the score should effectively stand in for and displace the work. He argues, "a score, as I conceive it, is a character in a notational language, the compliants of a score are typically performances, and the compliance-class is a work." (Goodman, 1976, 173) We return to this later.

### Working Notation

As a contrast to Goodman's abstract theory, Reynolds (2005) introduces some concerns regarding the use and function of musical notations from the point of view of a twentieth century composer, and with reference to his "creative continua" (introduced above). As we shall see, the cultural context into which Reynolds puts notation is not so much the language for constraining and disseminating the definition of a musical work as it is the material which displaces musical ideas and makes them malleable. He also puts a strong emphasis on the value of interpretation of notational evidence, particularly in musical performance.

Before examining his functions of notations, however, it's interesting to compare his ideas on the content and nature of notation to Goodman's theory. For Reynolds, written symbols (i.e. words) are subordinate in notational application to graphic symbols. Graphic symbols, he argues, have two principal functions: they either stand for events ("a generally agreed upon and well-defined sound") or for actions (a sign "that implies a certain physical action or chain of closely synchronized movements") and are most "suited to one-to-one correspondence with singular events". In order to represent "extended processes" or sequences of events, Reynolds argues that grouping mechanisms or "gradients" are applied to sequences of symbols. He also describes as a "common technique" (Reynolds, 2005, 136) the use of two-dimensional space to represent pitch against time.

Like Goodman, Reynolds describes the construction of symbols from collections of marks. As an example, he compares the construction of Latinate letters with that of Chinese ideograms, arguing that the component marks (or "radicals") of ideograms are meaningful in themselves and therefore correspond to Goodman's "prime inscriptions", whereas the component marks of letters are not. The meaning of an ideogram is derived from the combined meanings of its components, making it an example of Goodman's "composite inscriptions". Reynolds continues, the ideogram's "meaning resonates, even includes a certain graphically suggestive aura." (Reynolds, 2005, 137)

Returning to his functions of notations, Reynolds introduces what he calls the "most familiar notational function" of "transmission of intent", arguing that each stage of his continua "involve storage, internal or external, of items unconsciously or consciously coded." (Reynolds, 2005, 131) He goes on to describe "storage" as "the process by which the composer's intentions are made available for examination at some distance in time and space." (Reynolds, 2005, 132) (cf. Latour's "mobilisation"). Unlike Goodman's appeals to some idealised notion of the musical work and its perfect reflexive relationship with the score, Reynolds views the transmission and storage functions of notation as economically grounded, notation's earlier uses being essential in a time when musicians couldn't travel so easily and before the invention of sound recording. He goes as far as to describe what he calls a current "deemphasization" of notation as a result of technological developments allowing wider communication and better documentation of musical events. Similarly, as has probably been the case in the past, such technological developments lead to a greater variety of musical works which "are not so easily trimmed and packaged for representation by neat sets of symbols." Further, Reynolds argues, contemporary performance practice is still unfamiliar enough, compared to the "professional standardization" of "more commonly practiced forms", to require the presence of the composer. Clearly, unlike Goodman, Reynolds does not value notation much as a transmission and storage medium, but rather as a compositional tool.

For instance, Reynolds emphasises the power of context in notational systems. He describes it as "the small-scale formalization of 'understood' information" (Reynolds, 2005, 139). Given a context supplying sufficient information, symbols are required to convey less meaning and so can be clearer and more succinct. (Examples from CWMN include key signatures and phrase marks.) This is, on a small scale, an example of the summarisation process Latour describes as displacement: the context-defining inscription displaces some representational aspect of the inscriptions it subsumes.

Reynolds describes "abstract representation" (Reynolds, 2005, 132) as a method of organising the information or "materials" that a composer must work with, particularly in reference to the "idea" and "crystallization" stages of his creative continuum. He argues that the mental storage of ideas ("entities - ideas, images, effects") only becomes effective once those ideas have been "encoded and 'rehearsed' properly." Having more information available to the mind in "convenient forms", he argues, allows it to make more connections. "Once a valued experience has been grasped in some way, it is to our advantage to find an effective external notation for it." He also argues, however, that "symbolic representation is ... not just mnemonic" but also "allows us to rearrange, combine, order, and modify ideas or conditions in our own time to a degree and with an ease otherwise totally unmanageable." Again, Reynolds is describing exactly the same sorts of affordances that Latour ascribes to inscriptions under the labels "recombinacy" and "superimposition".

Similarly, Reynolds describes a function of notation which he calls "communication with the self", or the "usefulness of notation in the private world". His argument is essentially that notations are a good way of keeping track of ideas and being able to develop them in what he describes as a fast paced society. "Selective storage of experiences or viewpoints can provide valuable counterbalances to an otherwise unidirectional motion almost devoid of enrichment of feedback and *re*consideration." (Reynolds, 2005, 135)

Reynolds describes how notation used in performance can act as what he calls a "liberating reference" (Reynolds, 2005, 133). He argues that memorisation of music by performers is inadvisable and that notation should be available as a reference so that performers can concentrate on "more significant dimensions" (134), by which he means, for example, "tempo, color, rhythms and intonation". It is therefore important that "visual materials ... should be clearly fixed and easily discriminated" and other information about performance (such as "techniques and tasks"), he argues, is be better represented elsewhere in the creative continuum, and using "different notational modes". Goodman, on the other hand, makes no such provision for features of a work not being fully accounted for by the score, arguing that "[g]iven the notational system and a performance of a score, the score is recoverable." (Goodman, 1976, 178)

Reynolds describes how notation in Western music history has been "an extraordinary vehicle for the *generation of richness*" (Reynolds, 2005, 134) not only as a compositional tool, but also in its interpretation by performers. Particularly, he subscribes (like Goodman) to the notion of the musical work as a conceptual object independent of the score, and for which, he argues (*un*like Goodman) "[t]he printed document is a respected aid but hardly capable of total specification." He argues that this conceptual object, in its original form, "remains essentially unknowable", despite any attempts to inscribe it, and that a lot of the tradition of music in the West has revolved around musicians applying "substantial intellectual and emotional resource" in interpreting the notational evidence of those works.

He describes how the meaning of notational symbols changes over time, arguing that "actual practice", which is often different from the "reverence for urtext testaments", "negates the idea that the substance of the composer's intent is always inherent in the score's printed symbols" (Reynolds, 2005, 128). This, he argues, is because notation is always taking on new ideas from its technical and social contexts. He asks, if we can play early music using different instruments, tunings, etc., what do notational symbols really convey?

### Music Analysis

Also under the category of musical notations come numerous notations designed for analytical purposes. Such notations range from using Roman numerals to denote active chords in harmonic progressions (conforming to a pseudo-authoritative and tacitly culturally contingent theory of the construction of Western music harmony), through Lerdahl and Jackendoff's tree diagrams representing constructs such as their time-span reductions derived from their (much more explicit) theory of a generative grammar for musical cognition, to Schenker's notations (adapted from CWMN) for his prolongations which derive from his particular theories about how eighteenth century Viennese music was constructed.

These notations all demonstrate some important properties of inscriptions described by Latour. As we saw above, he describes a process of "cascading" of inscriptions. He argues that there is a maximum limit to the number of inscriptions someone can reasonably expect to be able to deal with. As a result, summarising inscriptions are devised which reduce the mass of inscriptions to a more manageable (though much less detailed) set of inscriptions in a continuation of the process of abstraction which saw the real-world objects made into inscriptions in the first place. As Latour describes it, "[s]omething has to be done with the inscriptions which is similar to what the inscriptions do to the 'things', so that at the end a few elements can manipulate all the other[s] on a vast scale. The same deflating strategy we used to show how 'things' were turned into paper, can show how paper is turned into *less* paper." (Latour, 1990, 21)

So Roman numerals, Lerdahl and Jackendoff's reduction trees, and Schenker's prolongation lines are each doing this work Latour describes of being a manageable (but still malleable) summarising inscription of a more complex underlying inscription, and importantly, in these cases, standing in an orthogonal relationship with those underlying inscriptions. This relationship is vital for their role as efficient summarising inscriptions: it is only because of their "optical consistency" with the inscriptions they summarise (in each case, CWMN) that these inscriptions function as useful summaries at all. A non-orthogonal, uncorrelated inscription such as a stand-off description of the harmony in words would still fulfil the role of summarising the original inscription, but not to the same degree of efficiency as Roman numerals inscribed orthogonally with the score, for instance. We return to these ideas in a discussion of musical structure and we'll also see later how they can be extended to the level of collections.

#### Transcription

Many of the encoding practices we examine later are deployed in the work of re-inscribing existing (paper) musical inscriptions, or *transcription*. As an example of an inscriptive practice, this has strong echoes of Latour's "cascades", in which successive re-inscriptions are arrayed in order and the last one is considered to be somehow authoritative. At each re-inscription stage (and this is especially true of musical transcription) some information is lost or gained which, as we consider below in a discussion on musical works, has a considerable impact on the nature of the thing inscribed.

It's also interesting to note that transcription practice has a whole methodology dedicated to the study of the relationships between transcribed artefacts, *stemmatics*. In stemmatics, unlike Latour's linear cascades, the sources are seen as taking part in a web of relationships. The field even has its own inscriptive practice in the form of stemmatical diagrams.

So it seems musical notations don't merely fit into cultural contexts, but rather are very much a product of a cultural context. As Hugo Cole says, "[a]reas of interest in a musical culture are reflected in its notations." (Cole, 1974, 8)

### 1.3.2 Recording

We now turn our attention to another major class of musical inscription, that of sound recording. The basic principal of audio recording—that a sensitive membrane should be excited by vibrations and, using a connected inscriptive device, inscribe those vibrations onto a medium which moves over time—is simple and has remained unchanged since its invention in the nineteenth century.

Here perhaps more than in the preceding discussion of notation, there are overlaps with the following discussion on inscription and technology. Both notation and recording are technologised activities, but recording fits better the mechanical reproduction class of technologies most often associated with technologised music.

Gitelman (2006) deals extensively with the very early days of recorded sound and draws some important conclusions on how people perceive new inscriptive methods. She describes the changing meaning of the word "record", and how concepts "intuitive" to written records were commuted to audio records. For example, many early commercial music recordings were preceded with (paratextual) audio announcements including such information as the name of the performers, the title of work, etc. Interestingly, this practice is now returning in the contemporary phenomenon of "podcasting" (or, more accurately, "netcasting") in which audio *programmes* (schedules of audio content) are distributed via the Internet for later consumption. Like the early recordings Gitelman describes, and unlike the highly packaged phenomenon of commercial records, netcasts are disembodied enough to require such a descriptive announcement.

Another importantly new aspect of early audio recording was that, as Gitelman puts it, each "transaction" was "exactly and *automatically* the same" (Gitelman, 2006, 46).

Reynolds (2005), as a composer, takes a different view of musical recordings. "The value of historical accumulation, of recycling and adaptation, is inestimable, and its loss ... would be severely damaging to the size of musical experience available to us. For this reason the idea of recording difficult compositions to serve as models for other performance attempts strikes me as undesirable." So for Reynolds, recording is an inhibiter to the preferable practice of interpreting the notated representation of musical material. He does, however, admit of recordings as "useful to sample unfamiliar sound materials or demonstrate performance techniques, certainly, but not to reveal ... a particular mold for eventuation." (Reynolds, 2005, 134–135)

Reynolds also makes a comparison between the healthy variety of recordings of well known pieces compared to those of less known works. "Commercial recordings of more popular repertory pieces maintain a balance through multiple treatment, but audiences learning the work of young or infrequently played composers often fall prey to the impression that the single extant recording of a work *is* the work, rather than one manifestation of a larger ideal." (Reynolds, 2005, 135) Of course, this situation is the norm in popular music, where performers are valued over works (and are often also the composers) and so their rendition of the music is considered authoritative.

## 1.4 Inscription and Technology

All media are in some way involved with technology, as all inscriptive methods are inherently technologised. To make an inscription requires a technology and so there is an inevitable entanglement between the process of inscribing and the technologies employed. As Gitelman puts it, "one of the burdens of modernity seems to be the tendency to essentialize or grant agency to technology." (Gitelman, 2006, 2)

As we hinted earlier, what's important here is the fact that technology is moving in the direction of the digital and inscription technology is no exception. Latour, as we saw earlier, describes the "homogeneous treatment as binary units in and by computers" (Latour, 1990, 16) of inscriptions, and Gitelman describes a view in which "digital media are all converging toward some harmonious combination or global 'synergy', if not also toward some perfect reconciliation of 'man' and machine." (Gitelman, 2006, 92)

An interesting feature of the introduction of new technologies is their tendency to attempt to fit into the cultural context of the technologies they replace or enhance. For example, we interact with our computers using keyboards borrowed from typewriters. This seems clear given that new technologies always come about within the culture of existing ones, and that, arguably, users prefer familiar (technological) paradigms. As Gitelman puts it, "[t]he introduction of new media ... is never entirely revolutionary: new media are less points of epistemic rupture than they are socially embedded sites for the ongoing negotiation of meaning as such." (Gitelman, 2006, 6)

Gitelman also makes the point that the degree to which users are aware of a technology is indicative of that technology's newness. She uses the development of scientific instruments as an analogy for the acceptance of new media technologies: eventually, scientists come to use the instrument, "the way one looks through a telescope, without getting caught up in battles already won over whether and how it does its job." Further, she also argues that "social processes" of use of tools make the "supporting protocols (norms about how and where one uses it, but also standards like units of measurement) ... self-evident". A consequence of this integration of new technologies into our ways of working is that, when they break down, we suddenly become conscious of them again and of the questions which would have accompanied their early adoption.

Cook (2005) makes a similar argument about the lack of uptake of computational tools for doing musicology. He argues that intuitive interaction methods won't "relieve us from knowing what's going on under the bonnet". For Cook, it's vital that users have a basic understanding of the "principles of music representation" (Gitelman's "supporting protocols") in order to know what affordances the technologies make.

Finally, it's important to remember that technological changes have similar impacts whenever they occur. For example, in discussing the development of musical notation from a method of defining or controlling social practice to a tool for making musical material malleable, Cole says, "[a]s has so often happened in the history of technologies, inventions designed for use in one field have found their widest application in other, unrelated fields." (Cole, 1974, 10) This also speaks of the importance of cultural context for technologies: a technology will only be successful if it finds a cultural context for itself, but that success may still come about even if the context changes over time or isn't the same as the technology's designer intended.

## 1.5 Digital Inscription

The specific class of technologised inscription under scrutiny in this work is digital inscription—those practices which involve the digital computer. As we shall see, the base digital substrate is very simple and all applications involving it require layers of representational semantics, sort of pseudo-inscriptions—an entity masquerading as a inscription which in fact encapsulates a number of more fundamental inscriptions, in order to become even practical. First, however, we consider something of the nature of digitality.

### On Digitality

Nelson Goodman includes a discussion on the nature of digital systems in his theory of notation. For Goodman, the distinction between digital and analogue systems lies not in any corollary with digits and analogies, but rather with what he calls "density and differentiation" (terms we considered above). "A symbol *scheme* is analog if syntactically dense; a *system* is analog if syntactically and semantically dense." That is, if, for any two signifiers or signifieds of a symbol system, there is always another signifier or signified between them, than that system is analogue, and, in Goodman's terms "undifferentiated in the extreme" (Goodman, 1976, 160).

Recalling Goodman's distinction between *schemes* and *systems*, a symbol *scheme* is digital if its characters (or signifiers) are "discontinuous" (discrete) and a symbol *system* is digital if both its characters and its "compliance-classes" (signifieds) are discontinuous. For Goodman, then, a digital system, as long as it is also "unambiguous and syntactically and semantically disjoint" (Goodman, 1976, 161), is notational.

Goodman also indicates some important properties of digital technology. He argues that while analogue instruments are best suited to "registering absolute position in a continuum" and "offer greater sensitivity and flexibility", digital instruments, just like notations, excel at "definiteness and repeatability" (Goodman, 1976, 161). Similarly, Gitelman emphasises the importantly new feature of nickel-in-the-slot phonographs that each "transaction" was "exactly and *automatically* the same" (Gitelman, 2006, 46).

Chandler expresses a similar attitude towards digital media, arguing that the difference between digital and analogue is a fundamental part of human experience and that we tend to regard analogue as more natural and digital as artificial. Like Goodman's digital instruments, he argues that digitality is equated with precision, "[s]ignifying systems impose digital order on what we often experience as a dynamic and seamless flux." (Chandler, 2002, 46) Again, the digital has the power to displace our inherently analogue world into discrete, comprehendable pieces.

Negroponte argues for a more literal view of the displacing power of digitality. His stance is essentially economical and he uses a distinction between "atoms" and "bits" as his principal metaphor. He describes a "bit" as "the smallest atomic element in the DNA of information." He continues, "[b]its have always been the underlying particle of digital computing, but over the past twenty-five years [up to 1995] we have greatly expanded our binary vocabulary to include much more than just numbers. We have been able to digitise more and more types of information, like audio and video, rendering them into a similar reduction of 1s and 0s." (Negroponte, 1995, 13) Negroponte's displacement, then, is much more physical in nature, "bits" displace "atoms" in numerous situations. His examples are based around the "information and entertainment industries" in which, he argues, "bits and atoms are often confused." He asks, for example, whether book publishers are dealing in bits or atoms? And gives the practice of video rental as an example where moving bits around (as we do now with, for example, on demand and Internet television) is much more viable than moving atoms around (as was more common in 1995 with physical borrowing of video cassettes).

### Advantages of Digital Inscriptions

We return now to Latour and his advantages of inscriptions. As we saw earlier, Latour's concept of the cascading of inscriptions and their becoming more homogeneous is "greatly facilitated by their homogeneous treatment as binary units in and by computers." (Latour, 1990, 16) So now we consider how each of Latour's advantages relate to the digital medium.

*Mobilisation*: the digital medium (particularly digital networks) greatly accelerates the mobilisation of inscriptions.

*Immutability*: digital inscriptions are much more mutable than paper inscriptions. Although both can be lost, it's much easier to erase or override (replace) a digital inscription than a paper one.

*Flatness*: paper inscriptions are inherently flatter than digital inscriptions. Digital methods like *random access* allow almost instantaneous traversal of data while paradigms such as hypertext not only defy flatness, but actually add a great deal of depth to digital inscriptions by allowing meaning to be "hidden" behind hyperlinks.

*Modification of scale*: considering the question of how far digital inscriptions enable the modification of scale requires dividing those inscriptions into two groups. The first group is what the digital graphics community call *rasterised* images, equivalent to digitally encoded waveform audio data. The inscribed data consists of numerous discrete samples of a continuous phenomenon such as image (continuous in spatial dimensions) or sound (continuous in the time dimension) encoded sequentially on the digital medium. The second group, on the other hand, are the methods in which a description of the inscribed object or instructions for how to render it digitally are stored on the digital medium. Examples include *vector graphics* in which the shape of an object is encoded as a series of co-ordinates and MIDI which consists of instructions for digital instruments. The first group are quite unsuitable for modifying scale as the computer is required either to discard inscribed information if the scale is decreased or to invent new information to fill in the gaps if the scale is increased[6]. Digitally stored instructions, on the other hand, can be rendered at any scale required with relative ease and suffer repeated rescaling without any loss of information. This distinction between sampled and rendered representations correlates to Goodman's distinction between syntactically/semantically dense and syntactically/semantically differentiated respectively.

*Reproducibility*: computers are perhaps the most successful replication machines yet devised. Because the most fundamental storage substrate for digital information is discrete rather than continuous, it's possible to copy every single element of it without having to lose information through summarisation or approximation. Furthermore, many copying operations carried out by computers involve built-in checks in which the information is subjected to a calculation to produce a *checksum*. This checksum is sent along with the information to the recipient which performs *the same calculation* on the received data. If the checksum it calculates matches the checksum it received then it is very likely that the information was received without error.

*Recombinability*: Latour describes this attribute of inscriptions as a consequence of their mobility, immutability, flatness, scalability and reproducibility. Therefore, any assessment of the suitability of digital inscriptions for recombination will depend on the extent to which digital inscription fulfils those criteria. This, of course, has been the topic of discussion so far. Without repeating too much of the preceding discussion, it should be clear that digital inscriptions are good candidates for recombination as a result of their *mutability* and reproducibility, it's easy to alter a digital document and easy to copy digital information from one source into another. However, digital inscriptions which don't scale well (as described above) will not so easily suit recombination as, using Latour's phrase, they lack "optical consistency" (Latour, 1990, 19), any differences in scale make them incompatible. Negroponte also celebrates this property of digitality, saying, "bits commingle effortlessly ... and can be used and reused together or separately." (Negroponte, 1995, 18) He goes on to describe this as the basis for "*multimedia*".

*Superimposition*: From the point of view of digital inscriptions, superimposition has much the same dependency as recombination. As long as two inscriptions are sufficiently flat and are scalable, they can be superimposed. Some important advantages that the

[6]That's not to say, however, that such processes are impossible. In fact, the better algorithms can enlarge images with moderate success and lengthen or shorten audio samples with only minimal distortion. My point, however, is that information is always and inevitably lost during these processes.

digital platform brings to superimposition, though, include interactive, transitory, and automatic superimposition. Because digital superimposition can result in a new inscription which leaves the source inscriptions intact, superimposition can be transitory; the source inscriptions are still available and can easily be retrieved. Similarly, the resultant superimposition can be continually re-generated from the source inscriptions allowing the superimposition process to be interactive (for example mixing audio, or morphing images). And the tools that process digital inscriptions, computers, are particularly good at matching patterns. Digital inscriptions, especially those containing textual information, may therefore be superimposed automatically by the computer by correlating their contents using pattern matching techniques.

Latour describes how inscriptions can, "after only a little cleaning up" be "*made part of a written text*". In the digital paradigm, this advantage is available as a result of the same processes that allow recombination and superimposition. Latour again describes the importance of "optical consistency" and "semiotic homogeneity" (Latour, 1990, 20) of the text and the inscription. On the digital medium, while the consistency is not strictly optical, all inscriptions are fundamentally made up of the same stuff, of bits and bytes. At a high higher level, many applications for document preparation allow the integration of data embodying different inscriptive metaphors (such as text, images, and tables) within one document.

Latour's last advantage of inscriptions—that they can be treated as geometrical equivalents of the things they represent, can be "merged" with "geometry"—is borne out digitally in the malleability of the class of digital inscriptions described above in relation to scalability: those inscriptions which are rendering instructions. Such inscriptions encode a dimensional correspondence with the objects they inscribe which allows them to, as Latour describes, stand in place of those objects for the purpose of measuring. Similarly, software designed to render those inscriptions often includes an interactive element which fulfils Latour's description of manipulability.

Negroponte describes some other important advantages of digital inscriptions. He introduces the concept of "a bit that tells you about other bits" (Negroponte, 1995, 18), or a "'header'", a working example of Latour's recombinancy and of cascading. The homogeneous binary stream can contain portions mixed up within the content which describe what the content is (how to interpret it, what it's about, etc.) but which are encoded in the same way as the content, as binary bits. (We shall see similar advantages in representation methods such as XML later.)

Negroponte also describes advantages of digital broadcast over analogue broadcast which are involved with the technological practice of employing more sophisticated devices at the transmission and reception points of digital broadcast than in analogue broadcast. Such devices may consume vast quantities of digital information at rates much faster than real-time and then are able to perform operations such as error correction (by manipulating the bits) and allowing consumers to select content (by storing and labelling the bits).

**Types of Digital Inscription**

Huron identifies four explicitly computer-based types of signifiers for music representation: "binary, alphanumeric, graphic, and sonic" (Huron, 1992, 7). We discuss binary and graphic signifiers as binary inscriptions and alphanumeric signifiers as text and XML inscriptions next, but Huron's observation of sound as a class of signifier is worth taking a moment to consider. As Huron admits, it's more natural to assume that sound is a class of signified in music representation, but he describes some situations in which sound can behave as a signifier. For example, many computer user interaction environments employ sound to indicate certain events or conditions, and it's not uncommon, as Huron explains, to use vocalised representations of musical objects such as "dut dut dut Dah" to indicate the motif from Beethoven's *Symphony No. 5*. However, we do not consider any further applications of sonic signifiers in digital music encoding practice.

### 1.5.1 Text

Although computers are fundamentally number processing devices, we most often employ them as text processing and communication devices. That communication may be either sending information over digital networks or committing it to paper for dissemination. Further, we normally get information into our computers using an array of keys inscribed with letters from the alphabet. As a result, we often perceive the most fundamental abstractions in computers to be textual, consisting of letters and numerals. Computers deal with these symbols by associating them with numerical values. This practice is standardised to aid in the compatibility of digitally encoded information between systems. Standards of text encodings include the American Standard Code of Information Interchange (ASCII) and UNICODE.

Following this, Huron describes the importance of what he calls "alphanumeric" signifiers in music representation practice, signifiers "associated with printed text in Western cultures" (Huron, 1992, 8). Many of the practices we examine later on take advantage of the correlation between musical pitch class names in the English, American and German notational systems (for example) employing letters from the Latin alphabet and the use of letters in computers. So in order to represent the pitch class C, for example, a computer file could contain the character `C`.

A common trope in discussions of text-based digital inscription methods is that of *human readability*. Such inscriptions are valued because, unlike the binary inscriptions we examine below, it's possible for a human user to open a text-based inscription in a text editor and make sense of its contents. As we shall see, this is especially relevant for inscription methods built on top of textual encodings such as XML.

Another property of text files often exploited in music representation is their row/column two-dimensional nature. Because character encoding systems such as ASCII include a special character which, instead of being rendering as a glyph, actually means break the current horizontal flow of characters at this point, it's possible to arrange text in computers on multiple lines. Thus characters can be vertically aligned as well as horizontally aligned. As we shall see in some of the music encoding practices below, this two-dimensional arrangement can be used to replicate dimensions of musical information. However, unlike CWMN which encodes the dimensions of pitch and time in scores, text-based music encoding methods (because they often encode pitch using phonetic rather than diastemic symbols) use dimensions such as time and instrumental part or harmonic movement.

Digital text inscription techniques have been applied in music representation through the proposal to extend the UNICODE character encoding standard with a collection of symbols used in CWMN. The proposal was authored by Perry Roland and is documented in the standard ISO/IEC 10646[7].

Other text-based inscriptions employed by methods examined later include the concept of tuples: simple lists of encoded values conforming to some predetermined record structure such as is found in databases. Examples include standards such as MARC (MAchine-Readable Cataloging) which allows the encoding of records in bibliographic

[7] http://www.lib.virginia.edu/artsandmedia/dmmc/Music/UnicodeMusic/ Accessed 10 May 2009

databases using simple symbolic codes to describe the content of each field. Similarly, the Resource Description Framework (which we cover in detail when we look at the Music Ontology project) allows the encoding of object-relationship-subject associative records called "triples" using a simple text-based syntax called N3.

### 1.5.2 XML

One of the key digital inscription or data representation technologies which is increasingly finding itself applied in musical contexts is the eXtensible Markup Language (XML). XML was standardised by the World Wide Web Consortium (W3C) in 1998[8] and was derived from the Standard Generalised Markup Language (SGML, see Goldfarb (1984)). It is intended to provide a mechanism for embedding semantic markers into documents but, importantly, provides no pre-defined set of markers. Each application of XML will define its own, domain-specific semantics within the syntactical constraints that XML enforces. For this reason it is often described as a "meta language" (see for example Dale and Lewis, 2006, 519).

#### XML Syntax

XML's syntax comprises *tags* which delimit portions of the target data, thereby designating those portions to have a particular meaning. That meaning should be expressed in the name of the tag[9]. The syntax rules ensure that tags properly delimit by requiring that each *opening tag* has a corresponding *closing tag*. The assemblage of opening tag – content – closing tag is called an *element*. Figure 1.1 shows a simple XML element. The content portion of an XML element, as well as containing text, may include further XML elements; a concept known as *nesting*. However, XML enforces a strict tree-like structure in documents by disallowing the overlapping of elements (see Figure 1.2). Thus the elements of an XML document are arranged in a hierarchy in which each element must have exactly one parent element. (We discuss some implications of XML's tree structure later.)

There are two other components of XML's syntax relevant to this study. The first is the *attribute*. Attributes consist of a *name* and a *value* (see figure 1.3) and may be added

[8] http://www.w3.org/TR/1998/REC-xml-19980210 Accessed 10 May 2009. Note that the specification has been superseded several times.

[9] It is important to note, however, that XML itself enforces no restrictions to make tag names meaningful. It is up to document authors to ensure that the semantics of their documents are clear by choosing appropriate tag names.

```
<paragraph>
  Lorem ipsum dolor sit amet ipso facto
</paragraph>
```

Figure 1.1: An example XML element

```
<a> <b> </a> </b>
```

Figure 1.2: XML elements may not overlap

to XML elements in order to alter or specialise their semantics. For example, the semantics of the tag name of an element may imply quite a general class (such as “person”) and an attribute (such as “gender”) could be used to make that element’s semantics more specific (such as “female”). The other syntax component is the *empty element* which is simply an element with no content. They have a special mode of inscription shown in figure 1.4.

### XML Idioms

Two principal and opposing idioms of XML markup are what Harold and Means (2004) call “narrative-like XML documents” and “record-like XML documents” (also commonly called “document-centric” and “data-centric”[10]). Narrative-like documents tend to employ much looser hierarchies, they often encode prose-like text, and they are the subject of numerous other idiomatic uses of XML (described below). Record-like documents, on the other hand, tend to employ much stricter hierarchies, they consist of numerous self-similar elements (like database records, or cards in a card index) and therefore involve much more structural repetition. In the discussions of applications of XML in music encoding which follow decisions as to whether music data are narrative- or record-like will be expanded upon.

Applications of XML in music encoding share one important concern with those in literature (particularly corpus linguistics) and any other field where XML is applied in

[10]There is an important terminological confusion here: although “document-centric” is common, it is also a common practice to refer to the disk files in which XML markup is stored as “documents” even when they contain record-like content.

```
<measure number="27">
```

Figure 1.3: Elements may have attributes

```
<note />
```

Figure 1.4: Elements may be empty

making transcriptions: reconciling XML's single hierarchy with often multiply hierarchic source documents. Documents are *multiply hierarchic*, when their structural or interpretative components don't nest neatly inside each other. The most common categories of conflicting hierarchies (drawing examples from both literature and music notation documents) are physical structures of documents (e.g. pages and lines or staves), syntactic, formal, or logical structures (e.g. sentences and paragraphs or bars and movements), and semantic structures (e.g. speech or phrases). Each of these categories can form logical and consistent tree-like hierarchies independently (e.g. each page of a document has exactly one parent document: the document itself), but those hierarchies regularly overlap with each other (e.g. it is perfectly legitimate and common for paragraphs or musical phrases to cross page boundaries). The syntax of XML will not allow these hierarchies to be encoded within a single document tree.

Several idiomatic uses of XML have been proposed and developed as a result of this problem (see DeRose, 2004, for an overview). In *stand-off markup*, first proposed (for XML at least) by Thompson and McKelvie (1997), the document encoder separates the hierarchies into their own disk files. The source document being marked up is left unaltered and tags denoting physical structure, formal structure, etc. are stored separately and refer back to the source document[11]. (Note the conflict between this practice and the fundamental principle in XML of *embedding* semantic markers within content.)

In *milestone markup*, the document encoder circumvents XML's single hierarchy by subverting its syntax rule which allows the use of *empty elements* (defined above). In this practice, empty elements are used to mark (or inscribe) the beginnings and endings of structures whose hierarchies conflict. So in place of the usual (and syntactically valid) element composed of an opening tag and a closing tag, the encoder inscribes an empty element where the opening tag should be and another empty element where the closing tag should be. These two elements' semantics reflect their functions as delimiters of a

[11] Interestingly, some SGML processors support a markup method called CONCUR which allows multiple hierarchies to be encoded within a single document by allowing the prefixing of hierarchy names to tag names. All elements of a given named hierarchy must obey the nesting syntax rules, but elements of different named hierarchies may overlap. No such feature is available in the XML specification.

```
<measure number="12">
  <phrase-start />
  <note pitch="C" duration="4" />
  <note pitch="B" duration="4" />
</measure>
<measure number="13">
  <note pitch="A" duration="8" />
  <note pitch="G" duration="8" />
  <note pitch="F" duration="4" />
  <phrase-end />
</measure>
```

Figure 1.5: The of use milestones to encode conflicting hierarchies

structural unit through, for example, tag names including "start" and "end" or attributes expressing similar properties (see figure 1.5). As part of the process, the document encoder often nominates one hierarchy as *principal* and encodes it using XML's supplied nesting syntax. By this action he arguably then designates all other hierarchies as somehow secondary or subordinate. (We return to this inscriptive use of XML elements below.)

Both the techniques described above involve subverting XML's syntax which, in turn, causes problems in dealing with the resulting documents. Tools built for processing XML encoded data (such as parsers) are generally optimised for use with syntactically normal documents and so any data which embodies practices other than these norms immediately sets up a requirement for manual or specialised handling of those encoding practices. For example, while XML's syntax (and by extension any XML parser) disallows elements without matching closing tags (it is considered an error and will result in the parser software refusing to parse the document), a structural unit which is delimited by two empty elements does not benefit from this built-in syntax checking. If the closing empty element is missing it is *semantic* error, but not a *syntactic* error and so the parser will not notice it. It becomes the responsibility of document encoders and maintainers to preserve the semantics of their documents without the aid of the technology. This requires a suitable level of expertise and attention to detail on the part of document encoders and subsequent maintainers.

For Wiggins (2009, 10) the single hierarchy problem is so severe (and its solutions so protracted) that he warns against employing XML in music applications at all. He accuses XML exponents of considering the concept of overlapping to be anathema to

```
<note pitch-class="D" octave="4" stem-dir="up" />
```

Figure 1.6: Empty elements are often used to encode non-textual entities

hierarchies, that they consider the tree to be only the class of directed graph which is hierarchical. Wiggins argues that the tree is only a sub-set of directed graphs, all of which are hierarchical and all, therefore, capable of and suitable for representing music. He does, however, make reference to (and allows as an exception) one important idiomatic use of XML as prescribed by its designers. XML was born out of a Web-centric culture (notice it was standardised by the same body which maintains numerous other Web standards) and, as such, was intended as a format for exchanging data between applications using the Internet, an *interchange format*. Wiggins argues that XML can be useful as an interchange format between music applications, as a kind of temporary or transitory inscription.

The final XML idiom described here is another subversion of XML's empty element syntax. Similar to the milestone idiom, some encoding practices make use of empty elements and attributes to represent objects and their properties without *marking up* any content. The elements are not being embedded as markers to indicate the semantics of some existing data, they are the data themselves (see figure 1.6). This is quite a common practice in music applications of XML because (as we shall see later), unlike text applications of XML, music has no pre-existing, un-marked up content to apply tags to.

### Defining XML Semantics

As well as the markup syntax described above, there are two other XML-related technologies referred to in this study which we briefly cover now: Document Type Definitions (DTD) and the XML Schema Language. Both are intended for the purpose of describing the required structure (which elements and attributes are allowed and in what order) of a class of XML documents. They allow some authoritative party to define or constrain the allowed semantics of compliant documents, acting somewhat like a template or blueprint or even (if the performative metaphor is not too stretched for this admittedly synchronic form) a score. An important social implication of these techniques is that the people who dictate the allowed semantics of documents and the people who inscribe those documents need not be the same people. The whole activity of defining encoding methods (here for music, but it applies in other domains) is predicated on these technological and social constructs.

### Inscribing XML

As well as the inscriptive nature of XML's components, issues of inscription as *process* arise in relation to XML. We hinted at some of these issues above (decisions made by document authors etc.), but we will now consider them in more detail. The terminological corollary of inscribing in XML practice is *marking up*; the work of applying semantic markup to a text. This summary immediately pushes the first issue to the fore. In much markup practice the content is a given, it exists before it is marked up and the marking up process is literally one of applying markup to that content, typing in the tags. Examples include the encoding of literature where the prose (or verse, etc.) exists as a body of text in which XML elements can be embedded. Similarly, more data-driven applications may draw content from databases and apply "record-like" markup to it. For music, however, the question of pre-existing content is more involved. Even if a piece of prose is not available in digital form, it can be transcribed from an existing (paper) inscription by taking advantage of the correlation between the glyphs which make up letters in written language and the glyphs most commonly processed by computers (those of ASCII and Unicode for example), glyphs which take the form of letters, numbers, white-space, punctuation marks, etc. the semantics of which is provided "free" by computer systems as we find them[12]. For example, computers are provided with an implementation of alphabetic order (the correct sequence of letters) which makes word sorting tasks trivial. The choice of these glyphs as fundamental in computer design reflects the most common concerns of those who designed and those who make use of computers: encoding and processing numerical and linguistic (especially in the early days of computing, *English*) content. As such, they make no provision for the glyphs required by (notational) music applications; it's not possible, as it is with literature texts, to *type music in* to a computer with any kind of direct mapping between typed symbols and musical symbols[13]. Furthermore, as discussed above, the grid-

[12]The digitisation of text is not always this simple, however. Often encoders are faced with glyphs which are difficult to interpret because of damage to a manuscript, for example, or with glyphs which don't have an easily available equivalent in the standard sets of characters made available by computers. In these cases, if XML markup is being applied, similar solutions to the techniques often employed in music markup applications can be used.

[13]Note that some score editing applications provide input methods which are computer keyboard-based for the sake of speed of data entry. These applications, however, don't implement the kind of one-to-one mapping between keystroke and musical glyph I am talking about. Each keystroke is interpreted by the program and results in (probably numerous) encoded elements.

like, two-dimensional structure of computer files does not provide a suitable container for diastemic music notation[14]. Like musical glyphs, the dimensions and structure of music notations documents can't be *typed in* to computers.

So the process of inscribing XML for music applications is often not one of marking up, but one of (as described above) subverting XML's syntax in order to encode—from scratch—the components of a musical document. As we shall see when we examine some encoding practices later, these uses of XML syntax include "record-like" representations of musical components (e.g. successive, self-similar, records encapsulating the properties of a note), empty elements functioning as inscribed objects (e.g. an empty element with the tag name "note" and including various attributes encoding the properties of that note), and container elements and *milestone* elements representing musical structures such as bars or phrases. These uses all emphasise the point made above that music XML encoding methods do not *mark up* existing content. For example, the "record-like" encodings often use sub-elements to encode properties of, in this case, notes. While these properties may be inscribed like `<pitch>A</pitch>`, the content of that sub-element ("A") did not exist before the element was inscribed, it was not *marked up* as a `<pitch>`.

Finally, the last issue we consider regarding the process of inscribing XML is the fact that it is an engaged-in process at all. In the case of digitised textual scholarship, marking up texts is much more of a manual task. Their communities' discussion forums (such as that of the Text Encoding Initiative[15]) are often taken up with questions such as discerning the semantics of a particular inscription and which available encoding element best describes it. The community of scholars engaged in hand-coding music using XML methods, however, is much smaller. As well as issues around the demand for and interest in producing digital transcriptions of musical texts, there are some more immediate factors pertaining specifically to the marking up of music. The first is to do with software tools of two varieties: music notation editors, and implementations of algorithms for automatically transcribing other digital music inscriptions into XML formats. The *MusicXML* format (described in detail below) has become increasingly popular as a interchange format between music notation editors and so many of these software packages now provide tools to generate *MusicXML*—to do the markup process—automatically.

[14]Below, we examine the **kern encoding method (part of the Humdrum toolkit) which does take advantage of this two-dimensional structure for encoding music.

[15]http://listserv.brown.edu/archives/tei-l.html Accessed 10 May 2009

### 1.5.3 Binary

The methods of digital inscription we have examined so far have been quite far removed from the most fundamental means by which computers encode and process information, that is *bits* (a contraction of binary unit, an atomic value which can be only either on or off, true or false, or 1 or 0) and *bytes* (groups of eight bits). All other digital inscription methods are built as layers of abstraction on top of this layer, with the text layer (described above) being one of those nearest the bottom layer. In fact, there's no technological reason why computers shouldn't provide standardised and universal music processing abstractions based upon their true-fundamental numerico-logical foundations, it's simply a matter of common cultural practice that they favour text in their most fundamental abstractions.

Huron describes the concept of binary signifiers in music representation applications, saying that they "provide the ultimate substrate upon which all other types of computer representations are based." (Huron, 1992, 7) He also uses the term "near-binary" to describe human *un*-readable digital inscriptions such as audio and MIDI data.

As we described above when we considered digital inscriptions in relation to Latour's advantages of inscriptions, many binary encodings are *sampled*, they consist of numerous snapshots of a continuous phenomenon encoded as a binary number which, when rendered either in quick succession (in the case of temporal phenomena such as sound) or sufficiently close together (in the case of spatial phenomena such as image) give the illusion of recreating the continuousness of the original.

As well as the obvious musical application of encoding sound as sampled data in a binary representation, there was considerable interest and optimism in the 1980s for encoding CWMN graphically and dispensing with any symbolic layer. "The arrival of the Macintosh platform, with its graphical user interface, in the early Eighties created a strange new environment for conceptualizing musical notation. Why represent the object at all? Why not simply draw it, store the image for future reference, and finalize the placement with a mouse?" (Selfridge-Field, 1997, 32) Huron describes how some scholars thought that digital graphical representations of music would render symbolic representations "fundamentally uninteresting" and that "in the representation of X as Y, the denotative mapping is an arbitrary tautology" (Huron, 1992, 6). He goes on, however, to argue that such graphical representations only displace the representational issue which, especially when considered by "applications-sophisticated users", repeatedly pushes itself to the fore. “The problem of representation refuses to go away merely by the introduction of graphics.”

# Chapter 2

# Representation

Having established something of the nature of what it means to make an inscription, how inscriptive practices relate to musical practice, and what role inscription plays in digital culture, we now go on to consider how inscriptions may be able to take on semantics or meaning, how they may become representational.

First, we return to Latour's concept of "mobilisation", the power of an inscription to displace a real-world referent. This displacement forms the basis of any claim an object may make to be a representation of something.

Winston describes a representation as "a set of syntactic and semantic conventions that make it possible to describe things" (Winston, 1984, 254). This division between syntax and semantics will become increasingly important, but initially, a corollary can be drawn between inscription as syntax and representation as semantics.

Brachman and Levesque employ a kind of spatial metaphor to describe what they call the "philosophically vexing" work of representation. A representation is "a relationship between two domains, where the first is meant to 'stand for' or take the place of the second" and in which the first domain is "more concrete" (Brachman and Levesque, 2004, 3). As well as this being a metaphor to convey the notion of representation, it is clear that the process described is very related to any working definition of "metaphor" itself.

Brachman and Levesque, in their work on knowledge representation in computational applications, concentrate on a specific class of "representor", what they call the "formal *symbol*", "a character or group of characters taken from a predetermined alphabet." Like Latour, they argue that symbols, and indeed all representations, are easier to "deal with" (Brachman and Levesque, 2004, 4) than the concepts they represent.

### The Work of Representing

Goodman attempts to address the question of what constitutes "realism" in representation, considering such concepts as resemblance, deception or illusion, and information content. He argues that the measure of the realism of a representation is not the quantity of information it conveys but "how easily it issues".

In a consideration of the differences between pictorial and verbal representation methods (between "representation in general and description"), Goodman concludes that representation is a "symbolic relationship that is relative and variable" rather than being a "mirroring", "physical" (Goodman, 1976, 42) process.

Like Brachman and Levesque, Goodman actually favours the representational power of the symbol over any other class of signifier. He describes how symbolic representations require little effort to comprehend and very unconscious interpretation. However, he does concede that "this depends on how stereotyped the mode of representation is, upon how commonplace the labels and their uses have become." (Goodman, 1976, 36) He goes on, "[r]ealism is relative, determined by the system of representation standard for a given person or culture at a given time. ... This relativity is obscured by our tendency to omit specifying a frame of reference when it is our own. ... Realistic representation, in brief, depends not upon imitation or illusion or information but upon inculcation" (Goodman, 1976, 37–38). As we shall see later, semiotic theory accords with this view, holding that the symbol is the most abstract of representations but gains its meaning through convention.

Brachman and Levesque describe some properties of dealing with symbolic representations which are reminiscent of Latour's inscriptions. They describe their field of knowledge representation as "the field of study concerned with using formal symbols to represent a collection of propositions believed by some putative agent." They argue that reasoning is an essential component of any working representation and describe it as "the formal manipulation of the symbols representing a collection of believed propositions to produce representations of new ones." They argue, like Latour, that the concrete nature of symbols, their manipulability is key here, you can "move them around, take them apart, copy them, string them together".

One important point they touch on is the inescapability of representations. They argue that manipulating symbols which represent propositions results in new sets of symbols which represent propositions. The process is entirely concerned with playing with symbols from start to finish. And this situation is in no way unique to digital knowledge representation. The notational, symbolic nature of practice in disciplines such as mathematics has been consciously addressed as far back as Leibniz.

Huron (1992, 9–10) opposes his own theory of representation against Goodman's theory of notation. He rests his theory on a particular interpretation of "signification" in which the semiotic concept of "signifiers" is prominent and the word "significant" comes to mean "worthy of being signified". He compares his "signification" with what he calls Goodman's "woolly" notion of "resemblance". However, a more careful reading of Goodman (as above) reveals that he treats resemblance as only a part of any notion of representation.

Huron, however, feels thus able to form the definition: "A signifier (X) is deemed to be a sufficient representation of a signified (Y) if and only if there is no significant property of Y which is not also implied by the signifier X." (where, as noted above, "significant" means "'worthy of being signified'"). As we shall see later, such constructs may be adopted to take into account the cultural practice of making representations, as in Marvin Minsky's "To an observer B, an object A* is a model of an object A to the extent that B can use A* to answer questions that interest him about A." (cited by McCarty, 2005, 25).

Huron (1992, 10) also argues that his signifiers don't actually represent themselves because they can only work within an appropriate context. Many of the properties of a signifier are drawn from its context (or sign system, in semiotics terms). The extreme conclusion he postulates is that in order to represent anything you need to give the whole world as a context within which to understand it. Using this, he makes the important point that in fact what's useful about a good representation is its ability to abstract away the irrelevant details of an object and its context and leave just the details which are pertinent to the task at hand.

Huron also extends this to a pragmatic point about the cultural practice of designing representations: "The essential point is that in order to represent something, its properties must be interpreted according to some proposed utility" (Huron, 1992, 10), i.e. you need to know what your representation is for.

Now we have established some notion of the nature of representation we go on to consider some more detailed theory around how representation works as well as some cultural practice of making and using representations.

## 2.1 Syntax, Semantics, and Semiotics

Much of the literature on representation emphasises the distinction (and relationships) between syntax and semantics. Beynon-Davies (1991, 40) describes syntax as the rules for "combining" and "arranging" symbols "to form statements in the representation formalism". Formalisms he describes as such digital representation paradigms as production rule systems, "structured objects", and predicate logic, while the semantics of a representation are how the constructed statements should be interpreted.

John Searle's (1980) Chinese Room analogy illustrates an important distinction between syntax and semantics. In this thought experiment a man known not to understand the Chinese language is put in a room with a look-up table mapping questions expressed in Chinese characters to suitable answers, also expressed in Chinese characters. He is then passed pieces of paper with questions written on them using Chinese characters and must use his look-up table to pass back out pieces of paper with the answers written on them in Chinese. From the point of view of the external questioner who understands Chinese writing but can't see what the man in the room is doing, it appears the man understands Chinese. Searle argues that many applications of semantics in computing are like the man in the Chinese Room, they are actually just clever syntactic manipulations which simulate rather than replicate whatever process humans employ to make meaning out of symbols.

The concepts of semiotics give us a means by which to describe how syntax and semantics function within sign systems such as languages and notations. Central to semiotic theory is the nature of the sign. Two principal models exist: those of Ferdinand de Saussure and of Charles Sanders Peirce.

The Saussurean model of the sign consists of: the *signified* (which is a concept) and the *signifier* (the thing which represents it — a sound pattern, image, etc.). They are purely psychological, "form rather than substance" (Chandler, 2002, 18). These two components together make up the *sign*. Saussure also held that signs only work as part of a sign system, meaning is constructed through structure or relationship of signs, not through reference, "signs refer primarily to each other" (Chandler, 2002, 22). Similarly, signs are meaningful because of their difference to other signs. For Saussure, the association of signifier with signified is arbitrary and dictated by convention. In language, Chandler (2002, 26) points out, even onomatopoeic words are different in different languages.

Peirce proposes a tripartite model of the sign: 1) representamen (form of the sign);

2) interpretant (the sense made of the sign); 3) object (to which the sign refers). This model includes the object itself (which Saussure's doesn't). Peirce seems to argue that the interpretant is somehow conceptually or psychologically active—it initiates a process of sense-making in the mind and results in a sign. Similarly, Umberto Eco describes a sort of signification chain forming. It's a diachronic process and, as such, is in opposition to Saussure's static, synchronic, purely structural model.

Chandler also presents a modern model of the "semiotic triangle": "sign vehicle [A] — sense [B] — referent [C]" (Chandler, 2002, 35), and notes that many media theorists also like to replace B with the user himself to emphasise process.

An important aspect of semiotic theory is how signs come to have their meaning. As we saw earlier, Peirce proposed a taxonomy of signs which includes iconic, indexical, and symbolic signs and both Peirce and later scholars believe that sign systems tend to evolve from iconic/indexical to symbolic over time. But critics of Saussure's almost structuralist stance argue that symbolic meaning—conventional meaning—rather than being fixed and static (or synchronic) is dynamic and subject to change over time as a result of social processes (Coward and Ellis, 1977).

Latour also makes the point that the meanings of signs are conventional, arguing that semiotics shouldn't be ascribed any "mysterious" power (Latour, 1990, 6). For Latour, sign systems are made concrete in inscriptions and inscriptions are just material things whose interpretation is a matter of convention.

So the sign's power to displace comes about as a result of its being symbolic, its relationship to the thing it represents is dictated by convention and that convention is subject to change. So any digital music encoding method (which must establish some signifier/signified relationship) will also be dependent upon the cultural context which sets up and alters those relationships.

## 2.2 Classification

An important part of representing is the way that objects to be represented are put into groups as a result of being given a particular label and how certain of their properties become more important depending on what group they are in or what label they have.

Goodman argues, "[i]f representing is a matter of classifying objects rather than of imitating them, of characterizing rather than of copying, it is not a matter of passive reporting." (Goodman, 1976, 31) For Goodman, objects don't have a specific, predetermined, and stable set of attributes which are always available for inspection. Rather, they can be a member of any number of groups and it is the particular current group membership which determines which attributes are prevalent. His most important conclusion is that in order to classify an object you have to choose ("prefer", 32) some of its attributes, "application of a label ... as often *effects* as it records a classification". In other words, for Goodman, the act of choosing some attributes of an object to prefer actually causes it to be in a particular class.

Bowker and Star (1999) describe this classification process as ubiquitous but "invisible". They argue that we are constantly making classifications but notice very few of them. They go on to propose some "simple" "idealistic" principles for classification, stressing that these principles do not hold in reality. "1) There are consistent, unique classificatory principles in operation. 2) The categories are mutually exclusive. 3) The system is complete." (Bowker and Star, 1999, 10–11)

Just as the assigning of signifiers in semiotics is a process determined by social and cultural context, so the classification of objects is determined by cultural context. Butterfield (2002, 338), in discussing the nature of the abstract concept of objects, makes the point that the classification of something as an object—and even as a specific class of object—is always a cultural thing. There are two important consequences of this. The first is that the concept *object* "entails above all a relationship to a perceiving subject". Things are nothing to themselves, it's only through human (culturally informed) comprehension that they become what they are. Second, that a different culture could classify a physical entity as a different class of object. (Butterfield gives the example of a calculator. Interestingly, he argues that cultures that wouldn't classify a calculator by its function may not classify two calculators as of the same class.)

This act of classifying, for our present purposes, is a generalisation of the cultural practice of making the musical fact something inscribable, of choosing its pertinent attributes, breaking it into meaningful units, putting it into a group, and assigning a (digital) signifier to it.

## 2.3 Modelling

We now turn our attention to the dynamics of displacement and examine some approaches to the work of modelling. For an overview, we return to Goodman. "Few terms are used in popular and scientific discourse more promiscuously than 'model'. A model is something to be admired or emulated, a pattern, a case in point, a type, a prototype, a specimen, a mock-up, a mathematical description—almost anything from a naked blonde to a quadratic equation—and may bear to what it models almost any relation of symbolization." (Goodman, 1976, 171)

Despite Goodman's observation that the term "model" is so multifaceted, there are some important aspects of models and modelling which we focus on here. In describing the process of discretising or digitising a phenomenon for the purpose of assigning it a notation, Goodman argues, "[h]ere as elsewhere the development and application of symbol systems is a dynamic process of analysis and organization" (Goodman, 1976, 163). Similarly, Willard McCarty argues that a "model attempt[s] to capture the dynamic, experiential aspects of a phenomenon rather than to freeze it into an ahistorical abstraction" (McCarty, 2005, 24). It is this idea of modelling as process, as a cultural practice which is important.

The opposition of modelling and model is a useful way into the discussion and is well rehearsed in the literature. Goodman, for example, opposes the "denotative" model and the "exemplary" modelling (Goodman, 1976, 172–173), Geertz describes the difference between the "model *of*" (made after the event, describing an existing thing) and the "model *for*" (intended as a plan or set of instructions for how to realise a thing) (Geertz, 1973, 93), and, similarly, McCarty describes a model as either "*a representation of something for the purposes of study*" or "*a design for realizing something new*".

Like both Latour and Huron, McCarty argues that models are by nature simplifications of the things they model and that this is where their power as a manipulable representation comes from. As Huron describes, a model is useful because it allows the modeller to abstract away any unnecessary or unhelpful context or details. McCarty describes the importance of being able to "play" with a model in place of the often complex and contextually entangled real-world objects they represent.

One of the most important components in modelling as a cultural practice is the person doing the modelling. McCarty describes a three-way relationship between modeller-model-modelled and argues that modelling is "fundamentally relational" (McCarty, 2005, 24).

Like Brachman and Levesque's conception of representation as the mapping of one domain onto another, for McCarty, modelling is about the relationship between the domains of modeller and modelled mediated by the model itself.

In this dynamic relationship, McCarty argues that the model provides two important affordances: "tractability" and "manipulability". He describes the tractability of models as "complete explicitness and absolute consistency", arguing that modelling is a trial and error process. Models will either work as expected, produce an inexplicable success or produce an inexplicable failure. Therefore, McCarty argues, "as a tool of research ..., modelling succeeds intellectually when it results in failure" (McCarty, 2005, 26). In other words, the process of making a model is more important—teaches you more about the thing you're modelling—than whatever outcome it results in.

Similarly, of the manipulability of models, McCarty argues that they are "temporary states in a process of coming to know ... rather than fixed structures of knowledge". In fact, he argues, once fixed knowledge has been obtained it no longer requires software representations and the model becomes redundant. Many of the practices we consider later on fail to take this idea into account and generally attempt to make a definitive and universal representation of a musical work; they value the model over the modelling.

McCarty draws an important distinction between models and representations. He argues that representations can only be *of* something and never (unlike models) *for* something. For McCarty, representations have a mimetic quality and also stand in a closed relationship with the thing they represent. Unlike modelling, there is no *act* of representing, representations are always complete whereas models *for* are always dynamic and incomplete.

Lelio Camilleri describes the importance of the process of modelling (which he calls methodology) in interdisciplinary practice, particularly with relation to musicology. He cites Ruwet (1977) on the importance of "a methodology that allows evaluation and falsification" and uses "explicit theoretical formulations" (models) which "facilitate the recognition of failures and inconsistencies," (Camilleri, 1992, 172) a stance very similar to McCarty's heuristic modelling.

Camilleri is very confident of the importance of psychological models of music and goes as far as calling for a re-interpretation of the field of music theory in terms of some psychological percepts which may be involved in music comprehension. He argues that computers have an important role to play in such work because, although there are numerous paradigms for encoding "musical knowledge", they are all linked by the fact that they can be "formulated as computational procedures, or algorithms." This, he argues, enables the "formalisation of propositions which seek to connect music theory and the theory of musical cognition" (Camilleri, 1992, 173).

However credible or not Camilleri's stance on the relationship between psychological and music theoretic models may be, the important consequence is that his "methodology" is essentially an example of modelling practice. It is by doing the modelling that we learn about music theory. Similarly, Cook argues that an important skill for musicologists to learn is how to "operationalise musicological problems" (Cook, 2005, 7).

McCarty gives a helpful summary of his attempt to establish a philosophy of modelling: "the manipulatory essence of modelling, with its connotation of embodied creation, physically or metaphorically; the mediating role and ternary relationship modelling establishes between knower and known; the directed, vector-like engagement of the enquirer's attention, *through* the model he or she has made *to* the object of study; and the model's consequent function as an artificial agent of perception and instrument of thought" (McCarty, 2005, 38). These summarise well the idea of modelling as process and models as tools for the work of coming to understand a phenomenon.

## 2.4 Representing Music

Having established something of the nature and process of representing, we now consider how the phenomenon of music suffers being re-framed in terms of abstract models and examine some of the practices of making representations of music. Babbitt (1965) (after Kessler) proposes three "domains" of music representation: "auditory" (mental representations), "acoustic" (sound, waveforms, spectra), and "graphemic" (scores, recordings). The practices examined in this work fall into the graphemic category, though what needs to be considered here is how those graphemic practices map onto the musical object being represented, how they displace the music.

Of course, this is partly tied up with the difficult question of "where is the music?" Wherever you attempt to pin the music down to (the score, the sound, the performer's gestures, the listener's perception, the composer's intention) it always slips away and pops up somewhere else, seeming there to be more real. However, as we discussed in chapter

1, this intangibleness is rendered unproblematic by making and manipulating inscriptions in place of the music. Following this, Huron notes that one of the principal problems in music representation is knowing/deciding what the "signified" in music is. It could be sound, notation, performance gestures, "perceptual/cognitive constructs", "formal structures". He describes identification of signifieds as "the most crucial aspect of the design of representations" (Huron, 1992, 9). Similarly, West et al. (1991) argue that the choice of music attributes to represent "must depend on the aims of a particular theoretical enterprise." (West et al., 1991, 5) The discussion here, then, will be focused on how music can be turned into something suitable for inscribing.

In the context of an introduction to the applications of music encoding in artificial intelligence research, Selfridge-Field makes an important point on how music may be considered a representable thing, saying, "[w]here those concerned with notation see objects in spatial positions and those concerned with sound hear events and timbres, researchers in artificial intelligence and related disciplines [quantitative and structural linguistics and cognitive psychology] look for grouping mechanisms that place objects into semantic clusters." (Selfridge-Field, 1997, 35) Following Brachman and Levesque, we hold that any credible representation must make reference, as Selfridge-Field describes, to some sort of semantics.

As well as supporting the argument for a representation requiring a semantics, Selfridge-Field also describes some key possible conceptions of domains of music which may be representable, particularly from an analytical standpoint, "Sound exists in time, notation exists in space, and analysis can be based on either or both, or on elements of the 'logical' work not represented in either one, or even on implied information (such as accent) experienced in performance (the 'gestural' domain of some commentaries)." (Selfridge-Field, 1997, 35)

Cook (2005) makes an important point about the power of musical notation to represent. He argues that "*musicologists are used to working with highly reduced data*" and that scores are "simplified" representations, to the extent that they "*symbolize* rather than *represent*" (Cook, 2005, 3) music. He goes on to note the importance for musicology of dealing with "full-sound data" in "the age of web-based multimedia" arguing, "[i]t will always be the business of musicology to understand sound in terms of the means of representation that define the culture" (i.e., if a culture uses music notation to describe sound, then musicologists must embrace that notation). "*[W]orking with fuller data will open up new areas of musicology.*"

Now we consider some of the principal components, properties, and conceptions of music which are picked upon as suitable for "grouping" into "semantic clusters" as candidates for the "signifieds" of digital music representation methods.

### 2.4.1 Score / Performance

The first is the division between music as a static, synchronic text and music as a dynamic, diachronic performed phenomenon. Wiggins et al. (1993, §2.1), in the context of defining categories for surveying music representation methods, address the question of what it is that systems are attempting to represent. They highlight the distinction between the "'score'" and the "'musical object'", arguing that the score "'precedes'" the musical object, is a set of instructions for the realization of the musical object.

Similarly, Cook makes the argument that, although the score does indeed provide a model for musical performance or analysis, in fact that score requires interpretation. Cook argues that there are "missing elements" which "have to be supplied by the performer or the musicologist". For example, performers rarely play rhythm exactly as notated and timbre and texture aren't directly notated at all. For Cook, "notation is only an approximation" (Cook, 2005, 3).

Describing common musicological attitudes to the score/performance divide, Cook argues that musicologists like the fact that, while "the sound is intangible", "the score is tangible". He argues that this reduced nature of musical information is problematic when, in methodologies such as set theory, it is treated as a "mathematical formula" and not as a "cultural object[s] dependant on contextual interpretation" (Cook, 2005, 3). He also notes that even in recent techniques such as analysing sound, the tendency is to reduce the signal to a level similar to that of scores.

In the context of analysing performances by aligning them with scores, Cook (2005, 5) argues that this technique neglects some important concepts in how "performers perform and listeners listen." He argues that music is not only considered "'vertically'" (as an audible phenomena correlated to a score), but also "'horizontally'" (related to other performances). For Cook, each performance is not an isolated realisation of a score, rather "performances are relational."

### 2.4.2 Discrete / Continuous

The next conception of music we consider is that of it as a continuous flow compared to that of it as a series of discrete events. Many representational methods (CWMN, for example) depend on describing music using atomic units (such as *notes*), but how far does music suffer such discretization?

Goodman describes how discrete symbol systems are applied to continuous phenomena. "If the subject-matter is antecedently atomized, we tend to adopt an articulate symbol scheme and a digital system. Or if we are predisposed to apply an available articulate symbol scheme to a previously undifferentiated field, we try to provide the symbols with differentiated compliance-classes by dividing, combining, deleting; the fractional quantities not registered by our meter tend to be disregarded, and the smallest units it discriminates to be taken as the atomic units of what is measured." (Goodman, 1976, 162–163) This is true of CWMN which doesn't extend its diastemic representation of pitch to such musical objects as *glissandi* or trills. Rather, as Goodman says, it treats their pitch content as "fractional" and therefore un-notatable. By extension, therefore, any digital music representation method which borrows the discretization employed by CWMN will inherit the inability to represent such fractional phenomena.

Byrd and Crawford (2002, 261) talk about the problem of splitting continuous signals up into discrete and meaningful units. In the context of information retrieval, they compare this discretization in music with that in text applications. They argue that, in computational applications, there is an important distinction between words and concepts: computers are only capable of dealing with words, not concepts. It is important to keep in mind, therefore, that any choice of discretization is arbitrary from a computational point of view. Computers deal only with the syntax of words, not with the semantics of concepts.

West et al. (1991) describe various levels of what they call musical "subunits". They argue that the "lowest level (notes, chords)" and "higher level (movements, stanzas)" subunits are easy to understand and justify, but that the middle-level units such as phrases and lines are the most contentious and their definition is subject to opinion and "musical aptitude and experience" (West et al., 1991, 3).

Butterfield, in attempting to define the "*microscopic*" musical object, argues that our perception of musical phrases as objects (in preference to hearing each note as an object) is an example of a class of perceptual phenomena (image schemata) known as the "source-path-goal image schema", and places it in opposition to the "container schema". A phrase begins at its *source* and moves along its *path* to its end *goal*. He also argues, however, that the source-path-goal schema can be considered a kind of container, the source and goal are "boundaries" which contain the path: "arrival at the goal reveals (and enacts) the path's 'containment'." (Butterfield, 2002, 353) Butterfield further describes this phenomenon as the source-path-goal schema being "mobilised" to a container schema once the "closure" of the phrase is apparent. At this point, the phrase *becomes* a musical object. Here, some of Latour's advantages of inscriptions can be applied to the problem of musical discretization. Although the "media" in question is actually just whatever media psychological percepts are inscribed on, if the percepts can be considered inscriptions then the source-path-goal (complex) becoming container (simpler) is similar to Latour's "cascades" and displacement.

Clearly, then, the problem of discretizing music is party a perceptual one and some objects which would syntactically be candidates for treatment as atoms (such as notes) may not be perceptually atomic at all, but are rather displaced as subatomic by a perceptually more prominent object.

### 2.4.3 Structure

West et al. (1991) argue that the concept of structure has become fundamental to most modern music theory and propose two simple definitions. First that structure is the way that "objects and events ... may be decomposed into subunits and the relationships between them", and second that "it is an attribute which objects and events have to differing degrees." (West et al., 1991, 2)

West et al. (1991), as was common in 1980s systematic musicology, favour psychological models of musical structure. Consequently, they argue that perceptual notions such as "belongingness, succession, repetition" partly account for the perception of structure in music, though interestingly, they also make the point that a lot of formal education in music theory revolves around leading students to perceive correct structural interpretations of music.

As we saw when considering discretization in music above, West et al. (1991) employ the notion of "musical subunits" in building their definitions of musical structure. Their first definition involves how these subunits may be related to each other. They describe the "*belongingness*" relationship (subsumption) as including "the *part-of* relationship" and "the *contains* relationship". They argue for the prominence of multi-layered structures in music and in this context introduce the "*structural importance*" relationship between "components and the units to which they belong". They argue that units are either "*structural*", meaning that they play a role in their containing structure, or "*non-structural*", meaning that they merely "embellish" their containing structure. (West et al., 1991, 4)

Similarly, Wiggins et al. (1993, §2.3) argue that musical data unusually requires that single musical objects can participate in multiple higher level structural objects simultaneously. They argue that any system which made this difficult would exhibit poor "structural generality" (a concept expanded upon in chapter 4).

West et al. (1991) also describe a class of relationships "pertaining to the *qualities* of the components" such as "*repetition*" (which they argue is central to almost any notion of musical structure), "*transformation*" (they describe as a kind of "incomplete repetition" and as one musical unit is derived from another), "*elaboration*" and "*simplification*" (are complementary processes which describe how a musical unit is derived by either the addition or removal of "non-structural" units, respectively), "*initiation*" (when a lower level unit is used as the beginning of a higher level unit), "*continuation*" (when subunits are arranged in succession within a higher level unit), and "*completion*" (when a lower level unit is used as the ending of a higher level unit).

Butterfield (2002) addresses some of the more fundamental perceptual principles which arguably underly those of West et al. (1991). He talks about the psychological process of category formation, in particular about our ability to understand basic-level objects (by being able to understand how we could interact with them) and the abstract concept of an object. He discusses Mark Johnson's idea of *image schemata*: a conceptual framework we use to process perceptual stimuli by seeing things such as "containers, paths, links, left-right, up-down and part-whole relationships". Butterfield argues that the containment and part-whole relationships are the most fundamental as they are necessary even for an understanding of the abstract "superordinate concept OBJECT."(Butterfield, 2002, 333–336)

Latour's concept of the merging of domains through inscription has some relevance here. He argues that "theories" are able to "hold" domains, to "remove" the observational evidence they are based upon "from isolation" (Latour, 1990, 21), by homogenising it all into one inscription. Similarly, many theories of musical structure are made concrete by the way they are represented on paper. For example harmonic progressions are often inscribed using Roman numerals to indicate the chord in operation and are most meaningful when they are presented orthogonally with notation for the music to which they apply. And larger scale Schenkerian prolongations are often inperceptable and would be meaningless without the re-scaled reduction diagrams used to inscribe them.

## 2.5 Digital Representation

Having established something of what it means to make a representation we now pick up the digital thread which we left at the point of defining how digital techniques may be employed in making inscriptions and establishing and conforming to rules of syntax. We investigate how digital systems may be able to do the work of representing.

Before we examine some of the ways in which computers may build representational abstractions, we first consider the status of the computer as a tool for modelling. One of the important aspects of digital representation is that in order to abstract some real-world concept into a form suitable for representation by computer it is always necessary to be able to express that concept in clear structural and logical terms, the kind of terms in which computers are able to deal with information. McCarty argues that "computing forces us to resolve" "demarcational issues" (McCarty, 2005, 30).

### 2.5.1 Symbol Systems

The general case of the computer's ability to represent is the symbol system. As Brachman and Levesque describe the formal properties of symbols and their associated semantics as representational systems, so we now consider the representational implications of symbol manipulation by computer.

Simon (1969) introduces the notion of the computer as a physical symbol manipulation machine. The context of his discussion is an attempt to define "artifacts", which he does by considering their inner systems, external systems (or environments), the interface between the two, and the way in which the artifact attempts to adapt to its environment.

His conception of symbol systems is as dynamic, diachronic things. Physical symbol manipulation machines produce "evolving collections of symbol structures" (Newell and

Simon, 1976) over time. The machines allow arrangement of symbols to be held and for manipulation tasks to be carried out upon them.

Simon sees these arrangements of symbols as models of the thing being represented and the work of the symbol system as modelling, "[s]ymbol structures can, and commonly do, serve as internal representations (e.g. 'mental images') of environments to which the symbol system is seeking to adapt." (Simon, 1969, 27)

Like Brachman and Levesque, Simon sees the symbol system as moving from one arrangement of symbols to another. However, the difference for Simon is that his symbol systems are embodied in a machine and so should be able to take input from the external environment and their final state should be able to act on or feedback to that environment. "[The system] must have a means for acquiring information from the external environment that can be encoded into internal symbols, as well as means for producing symbols that initiate action upon the environment." (Simon, 1969, 27) This is a role that digital representations can play, they can be both the means by which to "acquire information from the external environment" and also of "encoding" the final state of "internal symbols" of the system for feedback into that environment.

Brachman and Levesque (2004, 11–12) describe Simon's and Newell's stances as splitting knowledge representation onto two levels: the "knowledge level" (the logical, semantic level) and the "symbol level" (the implementational, computational level). They also describe their own two-level conception of knowledge representation: first that representations are made of symbols which observers can ascribe meaning to; and second that the manipulation of those symbols is how the knowledge reasoning process comes about (Brachman and Levesque, 2004, 6). The key difference is that Brachman and Levesque don't make any provision for automated attribution of meaning for symbols or for acquisition of symbols from the external environment.

Brachman and Levesque are more explicit about how symbol manipulations may be carried out, making a case for first order logic as a way of talking about symbol systems. They argue that (declarative) languages require three elements: syntax, semantics and "pragmatics" (being able to specify how "meaningful expressions in the language are to be used"). Again, they argue that first order logic can't claim to make precisely true statements about the world because its symbols are never mapped to the real world, only to interpretations that are useful in a particular context. The meaning is defined more in terms of the relationships between the symbols and particularly things like objects conforming to particular predicates, the meaning of the predicates is actually unimportant (Brachman and Levesque, 2004, 15–16).

So remembering Latour's displacement, we conclude that all these mechanisms are merely representational. Each simply pushes the real-world entity away, replacing it with symbols—be they conceptual or mechanical counters—and justifies this by arguing that these symbols are more malleable, just like Latour's inscriptions. Here this unstable framework of signifiers is being applied to an equally unstable *real-world* of music—itself a tangled collection of significations.

### 2.5.2 Digital Paradigms

Having examined the general case of symbol manipulation, we now consider a few of the more specific representational methods which have been employed in computational applications.

Most of these methods are concerned with defining what Abelson and Sussman (1996, §2.4) call "abstraction barriers", that is, a way of managing the complexity of digitally encoded data by hiding portions of it behind appropriately named interfaces and treating those portions as atomic units. Beynon-Davies (1991, 108) describes four principle methods of abstraction: "generalization, aggregation, classification and association."

He describes generalisation as "the process by which a higher-order object is formed by emphasising the similarities between a number of lower-level objects" ("an implementation of the `is_a` relationship"), aggregation as an implementation of the "`is_part_of`" relationship, classification as "a form of abstraction in which a number of objects are considered as instances of a higher-level object" (an "`is_instance_of`" relationship), and association as "a form of abstraction in which a relationship between member objects is considered a higher-level set object" (the "`is_member_of` relationship"). So generalisation and classification are roughly the inverse of each other. Similarly, Beynon-Davies draws a distinction between association and aggregation saying that aggregation is about defining the attributes of an object while association is about defining what are essentially lists with well-defined constraints (Beynon-Davies, 1991, 108–110).

Wiggins (2009) also emphasises the importance of abstraction barriers (or "boundaries"), saying that they are "a theoretical process which can elucidate the nature of our data and the operations we apply to it." He argues that it is important to make the abstractions domain specific ("specified at the level of concepts and language naturally used in the processes embodied in the program"), giving the example that music programs should prefer functions called, for example, "`transpose-pitch`" to "`add-integer`".

West et al. (1991) describe several digital representation paradigms with specific reference to their potential for application in music representation. Their "slot-filler notation" is based on Minsky's *frames*. Frames represent objects and have *slots* which represent their attributes and may contain values. West et al. (1991) argue that, as all the attributes are encoded, they are therefore each individually distinguishable and retrievable, including attributes such as succession and concurrence (encoded using attribute names such as "following", for example). As a consequence the order in which attributes are encoded is irrelevant.

West et al. (1991) also describe various shortcomings of the slot-filler representation such as its inherent redundancy (for example, using frames for notes and bars, given a bar's tempo and notes' durations, the notes' metrical values are derivable but may also be explicitly encoded). They note the method's lack of flexibility, particularly in encoding unknown or inaccurate data. They describe how the symbol (slot) names used are arbitrary and optimised for human understanding. This, they argue, means that the system is not self-contained, the meaning of the symbols has to be obtained from a source outside the system (West et al., 1991, 7–9).

West et al. (1991) describe what they call "temporally structured representations" as representing the succession and concurrence relationships between elements by encoding it as the ordering of the elements. (CWMN uses this concept: succession is represented as left to right text setting and concurrence is represented by vertical alignment.) They describe various list structures as ways of encoding this representation digitally. Their concerns, however, are with the difficulties of representing lists in computers in 1991 (such as static versus dynamically allocated arrays and the difficulties of random access to elements). They describe using indexes (stored locations in a list) representing "timelines" to solve the difficulties associated with accessing elements in a musically relevant manner. They also discuss the problem of representing "a continuous function like time in terms of successive data elements" and describe solutions such as using time as an index to the array of elements, and quantising the time dimension into "time slices"(West et al., 1991, 9–11).

West et al. (1991) describe generative techniques in music representation as being applied in practices such as "the generation of a 'surface pattern' of notes from some more compact representation" and "the generation of an interpretation of a surface structure in terms of a deeper underlying structure". As digital paradigms, these generative representations would be manifest as *procedural* representations, that is, they are encoded as method or algorithm rather than as static (or *declarative*) data.

West et al. (1991) give as a simple example of generative techniques the phenomenon of repetition in music. Repeated structures may be represented by a summarising atomic symbol (an example of an abstraction barrier) which is associated with some generative algorithm which knows how to expand the symbol into full-detail musical information.

West et al. (1991) describe "production systems and programs" which involve the encoding of *facts* and of *rules* which are used to infer new facts into a *knowledge base*, a process Brachman and Levesque describe as "*logical entailment*", "[t]he job of reasoning, according to this account, is to compute the entailments of a knowledge base." (Brachman and Levesque, 2004, 10). Beynon-Davies argues that rules are " primarily mechanisms for managing the 'information explosion' inherent in any attempt to represent reality. They are a more concise way of representing reality." (Beynon-Davies, 1991, 25)

West et al. (1991) speculate on the possibility of representing Lerdahl and Jackendoff's generative theory of tonal music (Lerdahl and Jackendoff, 1983) as a production-rule system. They describe how, although some of the rules in the theory are deterministic, others are probabilistic and so wouldn't suit implementation in a production-rule system as such systems are not suited to resolving the conflicts such rules would generate (West et al., 1991, 15–16).

We now return to the distinction between procedural and declarative knowledge. Beynon-Davies (1991, 25–26) describes how knowledge base systems treat procedural knowledge as stored data, a "declarative representation". He argues that this representation strategy is beneficial because it allows the knowledge to be independent of the processes which manipulate it (inference, etc.).

Similarly, Wiggins et al. (1993) argue that programs that *generate* (declarative) representations of musical objects are procedural representations of those musical objects. Therefore, they argue, it's theoretically possible to treat those programs as a representation (or as encapsulating an "*implicit*" representation). But they also argue that it's preferable to work with declarative representations and therefore that procedural representations should be interpreted and rendered as declarative representations where possible.

Declarative representations have some important parallels with Latour's advantages of inscriptions when compared to procedural representations. They are potentially more mobile than procedural representations in that a procedural representation may require some specific execution environment which may not be available when the representation is transported. They are a lot flatter than procedural representations; no parts are hidden. For this reason, they are much easier to combine.

We briefly consider the *connectionist* or neural network digital paradigm. Originally modelled on the physical functioning of the brain, neural networks have become a model for generating complex algorithmic-like behaviour independent of any connection with how actual neurons may or may not work. They consist of numerous nodes with weighted connections between them which represent the levels of excitation/inhibition between the nodes. There is an input layer of nodes and an output layer of nodes (similar to Simon's and Brachman and Levesque's symbol formations representing the external environment and alterations to that environment), and any number of layers between. The operation which the network performs is encoded within these intermediate layers of nodes.

West et al. (1991) argue that connetionist models for music representation and processing are quite problematic because of the temporal dimension of music. Such models could consume excitation data for musical events sequentially, requiring some form of quantisation and also a structural template for each event. They note, however, that musical processing requires that some portion of musical data representing more than a single instant be available at any given time (much like memory in humans). They suggest a possible solution to this using a first-in-first-out queue of several musical events which is presented to the network each time a new event joins the end of the queue West et al. (1991, 17–19).

### 2.5.3 Non-digitality and Re-inscription

Finally, we take a moment to make some observations about the suitability of various digital encoding methods.

Many of the practices under scrutiny here deal with digital representation as not much more than a mere re-inscription. They simply encode, in a "human readable" form, musical representations which are very closely related to the note-centric paradigm of CWMN. Further, this *readability* is often considered a significant merit of such encodings. From a cultural practice point of view, this has obvious benefits in that it should, in principle at least, make those encoding methods easier for users to adopt. However, it does represent only a minimal level of displacement.

What is often left unconsidered is the unparadigmatic digitality (or non-digitality) of these methods. The computer becomes simply a convenient input and storage device, its digitality left un-exploited. There is another side to this trade-off: encodings which don't make any attempt at human readability and which *do* attempt to exploit the digital paradigm more fully. The principal example is the connectionist models we've just discussed in which the internal representation is effectively incomprehensible. Instead of digitising some either pre-existing or carefully designed (and often paper-inscribable) representation of music, neural networks allow the computer to build its internal representation for itself.

This balance between concerns is indicative of the difficulties currently facing the computational musicology community. Uptake of computational methodologies will depend on scholars being able to encode enough data in a manner which is efficient (in input), comprehensive enough in scope for the task at hand, and which is verifiably accurate enough. On the other hand, the methodologies may not be fully exploited unless they are given representations which make the most interesting (and often hardest) problems tractable.

# Chapter 3

# Documents

For its definition of the word "document" the Oxford English Dictionary gives, "[s]omething written, inscribed, etc. which furnishes evidence or information upon any subject, as a manuscript, title-deed, tomb-stone, coin, picture, etc." Some of the key points to draw from this are that documents are inscriptive, they are the general case of an inscription carrying medium. Also that they are informative, somehow they are capable of referencing things (ideas, objects) external to themselves. We now consider the implications of these properties to musical documents and their digital corollaries, drawing on some of the issues in inscription and representation raised in the previous two chapters.

### Document Polysemy

The word "document" is polysemous; that is, there are two similar, but also significantly distinct, concepts which come under the name *document*. They are that the document can be both a *class* of things, and an *instance* of a class, for example Frank Harrison's *Music in Medieval Britain* is a document, but so is my printed and bound copy of it.

Nunberg (1979) discusses polysemy in linguistics in detail. He argues that there are numerous words whose sense alters depending on context. Unlike homonymy (in which two unrelated words sound or are spelt the same, for example "bank" meaning a financial institution and "bank" meaning the side of a river), polysemic terms refer to essentially the same thing but in different senses. He gives examples like the sense of "game" in, "the game is hard to learn" (where "game" denotes a set of rules) and, "the game lasted an hour" (where "game" denotes an activity), and of "radio" in, "we got the news by radio" (where "radio" denotes a medium) and, "the radio is broken" (where "radio" denotes an object) (Nunberg, 1979, 148). He goes on to argue that such polysemous terms are "indeterminable in principle", as fixing a definite sense of any such term is neither a matter of convention, nor can it be syntactically demonstrated. He also offers an argument as to why it's non-conventional: the senses of polysemous terms are always best defined in terms of each other. For example, and to return to the subject of documents, documents as inscriptive objects are intended to represent some contents, while documents as contents are intended to be inscribed. Neither sense is prime. (Nunberg, 1979, 172) And, as Latour describes, to do the work of displacing an idea into a document (content sense), but then not to inscribe that content into a document (inscriptive sense) is to waste the displacement.

Unlike the inscriptive, static document *instances*, document *classes* may change over time as a result of the processes of cumulative and branching re-instantiation (or re-inscription, or transcription) and thus the relationship between document instances and document classes is rarely ever fixed. Users of encoding methods are the actors who engage in these processes. Each time they make an instantiation of a document class, they are involved in an act of inscription which is influenced by or constrained within an inscription technology and results in a new document instance which, in turn, contributes to the changing definition of the document class[1].

The case of manuscripts is informative. A manuscript is in some sense a class with only one instance, it's unique. But that manuscript could be a copy and would therefore, at a different level of abstraction, be a member of a higher level class. And at a later date, a re-inscription may be made of that manuscript which employs a different technology to that which was used to create it, for example it may be printed or encoded digitally. Such a re-inscription may also be intended for quite different purposes than earlier inscriptions of the document, for example it may be a critical edition with references, or an online edition with hyperlinks.

### Functional Requirements for Bibliographic Records

The bibliographic community has had to deal with the problems around the relationships between document classes and document instances extensively. In 1998 the International

[1] It is important to bear in mind, however, that the classification of an inscription—a document instance—into a document class is a matter of interpretation and that that interpretation is usually executed by a different actor and at a different time.

Federation of Library Associations and Institutions published a report on what they called *the functional requirements of bibliographic records* (FRBR) (IFLA, 1998). In it, they define four categories: "work" ("a distinct intellectual or artistic creation"), "expression" ("the specific intellectual or artistic form that a work takes each time it is realized"), "manifestation" ("the physical embodiment of an expression of a work"), and "item" ("a single exemplar of a manifestation") (IFLA, 1998, §3.1.1).

So the process we described above of new instances contributing to the definition of the class becomes new *items* being derived from existing *manifestations* and contributing to the definition of the *work*. And similarly, a manuscript is a unique *manifestation* (i.e. there is a one-to-one mapping between *items* and *manifestations*), but its parent *expression* may have multiple *manifestations*.

There is also an important process of divergence of works, expressions, and manifestations which is related to this phenomenon of re-instantiation. In his study of the practices of bibliography, Dahlström (2006) identifies two key activities which are engaged in within the subfields he calls "reference bibliography" and "material bibliography": "*clustering*" and "*transposition*". His "clustering" refers to the organising of bibliographical evidence into conceptual hierarchies such as the FRBR model, while his "transposition" refers to the way in which content is transmitted from one document to another (he uses the terms "departure document" and "target document"). This transposition is a formalisation which describes the process through which re-instantiation occurs. As content is *transposed* from a *departure* manifestation to a *target* manifestation by any means other than mechanical reproduction, its form will be altered, either deliberately (e.g., by augmentation or annotation) or accidentally (e.g., by omission or misinterpretation) and eventually may become different to the extent that it represents a distinct manifestation. Similarly, manifestations may be commuted to new expressions, and even expressions to new works[2]. This process has strong echoes of Latour's cascading of inscriptions. The practices Dahlström (2006) describes are the more deliberate and less gradual changes which occur in the production of scholarly editions. Many of his concerns, therefore, overlap with those surrounding the production of digital editions which we deal with here.

[2] Again, as above, these concepts are dependent on the interpretation of the available material, a judgement which Dahlström argues has been tacitly (and probably wrongly) assumed to be value free.

### Documents as Containers

Having considered something of the nature of documents as entities, we now explore further some of their relationships to works and their affordances for expressing meaning and being representations.

As Nunberg described with reference to books, documents as inscriptive entities are containers for information. They are concentrations of artificial intellectual content, deliberately constructed (hence "artificial"), drawing together, like Latour's inscriptions, different times and places. There are two levels on which documents may do the work of representing. They may be representational in their own right as an object, or they may represent through the semiotic qualities of their content.

That first class of a document's representational power is manifest in its ability to participate in a referential chain (or, more realistically, a web) as either a re-inscription of a document which precedes it, or as a source for a document which follows it. This stemmatical process corresponds to the discussion above of divergence in works, expressions, and manifestations.

The second class draws much more upon the representational affordances of the component inscriptions of the document. So the ability of a document of language content to be representational in this sense relies on the representational power of the language itself, essentially a kind of propositional meaning. Similarly, musical documents rely on the representational power of the inscriptions they contain. Although content such as musical notation does not have the same power of propositional meaning as language, it is nevertheless meaningful to those who can use it as, for example, instructions for performance or data for analysis.

Documents also command representational power through their ability to preserve transient phenomena as immutable containers. Levy (2001) introduces the notion of documents "talking"; because they are immutable they tell us about the time from which they originated. Also, they tell us about the linguistic form of expression that was employed in their production. In some documents this information is important (such as religious and legal texts). There is also a corollary in musical texts. The precise form in which musical content (or text) was inscribed tells us something about what the people who chose to write it down felt was important to record.

A prime example of this is the medieval practice of *musica ficta*. Singers up until around 1600 were trained to think in hexachords, diatonic scales of six notes existing in three different species depending on their arrangement of tones and semitones. The hexachord in operation would change over the course of performance, but any necessary sharpening or flattening of pitches was usually not notated. Singers were expected to be following the hexachord in operation and to know when pitches should be sharpened or flattened. So the documents which come down to us don't include this inscriptive detail.

#### Documenting

This brings us to a discussion of the act and purpose of making documents. With regard to music, Hugo Cole argues that, "[a]ll earlier notations seem to have shared the common purpose of preserving and safeguarding the music of a culture, whether by defining a theoretical basis, establishing the course of ritual performance, or by transmitting performing skills or the rules of the musical game." (Cole, 1974, 9) Similarly, Latour (1990, 25–26) references McNeill (1992) arguing that documenting is often carried out by institutions in order to increase or maintain their power. McNeill argues that you gain power over people by having knowledge and that, therefore, the more knowledge you can gather together in the form of documents (McNeill uses the term "files"), the more powerful you become. This power is very much related to controlling attempts made by institutions such as the medieval church in producing authoritative books of chant.

Weinberger (2002) makes some interesting observations about the status of documents in digital contexts. He argues that the term was appropriated by early commercialisers of the paradigm of desktop computing in an attempt to express the general case of any atomic unit of self-contained content, and that, before this, the term was much more specifically associated with formal documents such as legal documents. As a result, the use of "documenting" has taken on quite specifically digital connotations. (Weinberger, 2002, 37–38)

## 3.1 Musical Works

Above we established the notion of the *work* as "a distinct intellectual or artistic creation"(IFLA, 1998, §3.1.1). We now consider some of the problems the notion of the musical work may pose in order to make a case for documents being able to be carriers of works.

The status of the musical work up until Joseph Kerman's attack on musical formalism (Kerman, 1985) was as an autonomous, synchronic, and often canonised object. The work was a matter of fact independent of any realisation and immune to questioning criticism. Butterfield (2002, 327) describes how the "[musical] object is typically seen today as 'reified', as a mere text devoid of musical sound, a lifeless artifact that stands in for and devalues a living artistic practice."

Since the early 1990s, however, this conception of the work has been largely discredited. McClary (1993) argues that works are more a product of their social context of creation and reception than they are abstract, formal constructions. To demonstrate, she applies a feminist reading of narratives surrounding the musical work and concludes that works are not "ideologically neutral, autonomous objects," but rather "representations or subversions of a patriarchal heterosexuality."

Grammit (1998) argues that McClary's model actually leaves the musical object intact, the social contexts whose narratives and vocabularies she analyses just work around it. "[McClary's] analytic practice accepts and reinforces a conceptual separation of musical work and social context that has much in common with the practice of formalist analysis she critiques" (Grammit, 1998, 24). He argues, instead, that autonomy is only appearance, brought about by the Marxist-inspired social relations which come together to produce that appearance. That is, musical works are the result of the actions of all the people involved in them—producers, consumers, critics, analysts. In this way, Grammit commutes the musical work from being a meaningful object to being an instance of practice. Although Grammit's view of the way that musical works come about has credibility in this current tail-end of the trend for postmodern approaches, it is only by making inscriptions (which Grammit may well argue constitute just another form of work-making practice) that any of the actions which *cause* the musical work may become manifest. As Latour argues, the actions displace intentions and ideas, but that displacement is wasted unless it is inscribed.

Small (1998) takes a similar stance to Grammit, but expresses his process-oriented approach as a use of language. He proposes the use of "music" as a verb—to music—and hence expressions such as "musicking". He argues that all musical practice revolves around the performance and all parties involved (including composers, performers, and audiences) are musicking in some way.

Butterfield (2002) argues that the motivation amongst these scholars for breaking down the notion of "being true or faithful to a work in representing it in performance" and instead constructing a "process-oriented understanding of music" was that such an understanding values the act of performance as much as (or even more than) the act of composition and so increases the scope for practices which may be examined under the label of "musicology". He describes such practices as "organised around the production of works—styles, in other words, that resist explicit textualisation and commodification altogether." (Butterfield, 2002, 329)

Butterfield (2002) accepts the importance of process-oriented musicology, but argues that it still depends on, and even reinforces, the idea of the musical work. He takes the "'work-concept'" from Goehr (1992) (the idea of the musical work as an imaginary thing) as a point of departure, arguing that it allows you to deal with the musical work as an object without having to worry about any of its individual instantiations, or "contexts in which it can be encountered", and to "speak meaningfully about the structure, form and content of a piece of music." (Butterfield, 2002, 330)

He argues that this "work-concept" already contains all of the process ideas proposed by the anti-work scholars and so there is no dichotomy between object and process. Furthermore, he describes how dispensing with the work-concept actually doesn't benefit process-oriented musicology at all, arguing that "the work-concept continues to regulate musical practice in both originary and derivative ways." (Butterfield, 2002, 330)

In his argument against a strict distinction between what he calls music as "process" and music as "product", Cook (2001) makes the argument that experiencing a musical performance leaves an impression of a coherent object, rather than a few disconnected fragments.

Butterfield (2002) takes up this concept of a musical *object*, arguing that there is an important distinction between it and musical works. To demonstrate, he gives the example that Reich's *Music for 18 Musicians* is easy to classify as a work, but difficult to experience as an object (because it's long and difficult to follow any driving temporal force in it). And that improvisation is often easy to classify as a musical object (it makes musical sense and is easy to experience and comprehend) but difficult to classify as a work. How, then, he asks "do we objectify musical phenomena, or temporal phenomena in general? What is the ontological nature of such objects?" (Butterfield, 2002, 332)

To answer this question, Butterfield (2002) turns to Lakoff and Johnson (1980, 10) and their concept of the "*ontological metaphor*", the metaphor by which we can treat as objects things that are not obviously objects (such as being able to talk about "inflation" as a thing). Citing Lakoff, he argues that we are able to use this ontological metaphor as part of our unconscious perceptual apparatus: "concept OBJECT is *functionally embedded*" (Lakoff, 1987, 13–14). In the style of Lakoff and Johnson, Butterfield proposes the ontological metaphor "SOUNDS ARE OBJECTS" (Butterfield, 2002, 339).

He goes on to posit a thesis on the nature of the "musical object". He argues that there are two species of musical object: the "*microscopic*" and the "*macroscopic*". He describes microscopic musical objects as "discrete" sounds which "may or may not suggest the formation of larger musical wholes." His "*macroscopic*" musical objects are roughly akin to musical works. However, he argues against the idea of a hierarchy of musical objects with the macroscopic at its top (in "the manner of an *Ursatz*"). He argues, rather, that they are perceptual objects that come about as a result of a gradual process of interpreting, an "emerging wholeness" (Butterfield, 2002, 349).

Smiraglia (2001), in his work on defining the status of the musical work in music information retrieval research, describes how it has been considered "too abstract or difficult to define empirically", to be utilised as a practical entity, and has been dismissed in favour of the "document" as a primary entity for music information retrieval applications (Smiraglia, 2001, §1). This notion correlates with Butterfield's "macroscopic" musical objects and is one that Smiraglia attempts to oppose.

He makes the case, instead, for works as "vehicles of communication". They convey what he calls an "intellectual sonic conception" which, he argues, provides a useful abstraction barrier to subsume both the *composed* work and other classes of musical work such as improvisations. Furthermore, he argues, in opposition to Butterfield, that works are "created deliberately", and that they are "containers" for representations of "the thoughts, data, syntheses, knowledge, art and artifice of their creators." (Smiraglia, 2001, §3) So, for Smiraglia, the work is not a psychological percept generated as the result of listening experience, but a consciously conceived and constructed artefact.

Smiraglia (2001) goes on to argue that his "vehicles of communication" idea allows works to participate in chains of music production in which one (or several) works may provide the basis of another work, communicating "across time and space to new consumers." He calls this "the social role of works". From our point of view, it's an important recognition of the vital work inscription plays in the musical work. Just as Latour describes the power of the inscription to bring multiple times and places into one place, so Smiraglia ascribes the same power to musical works.

Finally, it's worth noting that these issues of what a musical work is are a common problem in computational applications where the computer always requires things to be defined in clear and unambiguous terms. I have personal experience in dealing with these problems in two slightly different domains. The first was in a project to produce an online thematic catalogue of the works of Benjamin Britten in which I was responsible for designing the database schema. The database was required to model not only the relationships between works and their constituents (movements, acts, etc.) which we generalised as "sub-works", but also the various sources related to them such as manuscripts, published scores, recordings, etc. The model was made even more complex, however, in attempting to capture the phenomenon of a revision to a work, especially in cases where a revised work consisted of a different combination of sub-works. In another project, I was required to model works for the Sonic Arts Research Archive project. This required thinking carefully about the nature of the musical work and particularly trying to understand the attitude some contemporary practitioners take to it. Numerous "works" in the SARA database are more like contexts for performance than any notion of a fixed work. These works often consist of a title and a particular configuration of (often electronic) apparatus which the performer (and, usually, creator) uses to improvise. They become concrete through documentary evidence (e.g. recordings, photographs) of each individual instantiation. Some artists even draw a distinction between instantiations by giving them unique titles (e.g. Shigeto Wada's *fpeak* has the date of performance appended to the title for each instantiation and David Casal's *Sound spotting* is postfixed with an incrementing number each time it is performed.)

Returning to our main thread, we are left with two opposing conceptions of the musical work: one a psychological percept which comes about as the result of listening experiences, and the other a more materialist, creator-initiated container for musical content. Both, however, are dependent on inscription for their existence. In the case of Butterfield's psychological percepts, to displace sound into discrete, meaningful sound objects, and further to displace those objects into coherent musical works, but not to inscribe that displacement is a waste (following Latour). And, if we take Smiraglia's concept of the work, we deal with a material entity which can only exist in an inscriptive form. The musical document, then, will now be examined as a container for the inscription of the musical work.

## 3.2 Musical Documents

Having established the concepts of the musical work and the musical object, here we consider how those concepts are served by (physical) "documentary entities" (borrowing Smiraglia's term).

We live in a culture which privileges the *written* and affords it credence often above sensory experience to the contrary. This is particularly true in relation to abstraction. Latour argues that this bias comes from "training (often in schools) to manipulate written inscriptions, to array them in cascades and to believe the last one on the series more than any evidence to the contrary." (Latour, 1990, 23–24) Similarly to this, we have the concept of literate music cultures, those which write their music down and which have lead to a whole body of musical practice which is only made possible by inscription. The ability to inscribe music and disseminate musical documents influenced the development of music profoundly.

Although the concept of the *urtext* has become quite unpopular, contemporary Western culture still has some considerable confidence in its musical documents. Such documents are used as instructions for performance, data for analysis, as containers for scholarly discourse (critical editions and musical analyses, for example), and are increasingly multi-modal (that is, multiple modes of representation such as notation and audio may be simultaneously present in the same document). This represents, also, a changing attitude in what may be considered a musical *text*. A rough definition of the postmodern notion of a text could be anything that can be referred to, in which case documents such as recordings could come under the umbrella of "text".

In this section, then, we will try to establish musical documents as entities capable of being a representation of (manifestation of an expression of, in FRBR terms) a musical work and to asses their suitability for digital encoding.

We begin by returning to Goodman, who argues that much discussion about notations fails to address the nature of scores and often treats them as "mere tools", assuming that they serve no purpose after the performance (Goodman, 1976, 127).

In opposition to this, Goodman argues that the score serves as the "authoritative identification of a work from performance to performance" and that all the properties of scores and notations derive from this function. He also notes the "exciting", but, he argues, subordinate application of scores and notation in "facilitating transposition, comprehension, or even composition" (Goodman, 1976, 128), musically specific species of the kinds of benefits Latour ascribes inscriptions.

Goodman goes on to argue that a score defines the class of a work and that performances are instantiations of that work (or class), "marking off the performances that belong to the work from those that do not". He is quite strong in his belief that a performance either complies to the score or not, though he does concede that the test only need be "theoretically manifest", i.e. it's not necessary for it to be easy to say whether a performance complies to a score.

He also argues that scores operate in a kind of notation system of their own, almost as atomic characters. He uses this idea to reinforce his position that a score should admit of only one compliant performance. "Not only must a score uniquely determine the class of performances belonging to the work, but the score (as a class of copies or inscriptions that so define the work) must be uniquely determined, given a performance and the notational system." (Goodman, 1976, 130) That is, it must be possible, given a performance, to determine which score (or character) it is of.

However, for Goodman, the score is only incidentally a physical entity ("a class of copies or inscriptions"), even though he affords it great documentary power. Smiraglia, on the other hand, who defines "documentary entities" as "unique instance[s] of knowledge", argues that "each documentary entity has physical and intellectual properties." That is, it's both a physical *thing* and a container for some abstract content (Smiraglia, 2001).

Smiraglia (2001, §2) argues that the containment relationships that hold between his "physical" and "intellectual" properties of documentary entities is key to the use of those entities in music information retrieval applications. For Smiraglia, the physical properties *contain* the intellectual. He also gives some examples from a taxonomy of types of containment relationships between works and documents, including: "simultaneous derivations, successive derivations, translations, amplifications, extractions, adaptations, and performances."

Cook (2001) takes another view, drawing on disciplines such as theatre studies to argue for notion of scores as “scripts” (rather than texts) in order to address some of the problems which arise as a result of tensions between texts and performance arts. This stance makes the performance not a “reproduction of [a] text[s]”, but merely a “cultural practice prompted by scripts” which “dissolves any stable distinction between work and performance.” (Cook, 2001) Thus, Cook is able to argue against the notion of a single work with many instantiations in “vertical” relation to it, and instead proposes numerous “ontologically equivalent instantiations, all existing on the same ‘horizontal’ plane” (Cook, 2001).

Elsewhere, Cook (2005), in the context of a discussion of what he calls the data-poverty of musicology, argues that musicology is the study of scores and not music and has been since the nineteenth century when, he argues, musicologists saw themselves as “musical philologists”. He argues that these nationalist scholars saw their task as to document their musical heritage through reconstruction of music using scores and, as a consequence, they viewed music as a “form of writing, abstracted from its contexts of performance and consumption.” He goes on to argue that early musicologists’ conceptions of music may have been like that of literature, that they were dealing with a corpus of works. He describes this view of musical information as “very narrow”(Cook, 2005, 3).

However, Cook argues that literate music cultures (such as the Western culture) “involve a constant tension between what is heard on the one hand and what is constructed or manipulated through notation on the other: as a technology of the imagination, music writing stretches people’s powers of perception”. For Cook, then, scores are part of the culture of Western music, and not merely a “highly reduced” and “unsatisfactory representation”. Using notation is “an essential part of what it means to study the music *musicologically*.”(Cook, 2005, 4)

So both Goodman (the nominalist) and Smiraglia (the practitioner) agree on the document being able to contain the work. And Cook affirms the musical document’s central status in Western musical culture. We now go on to consider in what ways these musical documents may become physically manifest.

## 3.3 Musical Media

Here we take our musical document, itself an inscriptive and representative (but still conceptual) manifestation of a musical work and consider how it may take on physical form by being inscribed on a medium.

As a substrate for inscription, many digital media are serial in nature and therefore suit well trace- or signal-like representations of music. As music in performance is temporal, such a linear representation as signal for performance is functionally adequate. This linear or serial arrangement of data is also generally translated into digital media for capturing notational representations of music as well. Although this is digitally sound, it fails to take into account some important non-serial properties of notational documents, such as vertical, backwards, and displaced relationships between elements. We examine some specific instances of practice later and find that this disjuncture between serial digital media and non-serial musical data is still a largely unresolved issue.

Considering media as culturally contingent entities, Gitelman makes some important observations about their physicality. "Like other media, inscriptive media represent, but the representations they entail and circulate are crucially material as well as semiotic. Unlike radio signals, for instance, inscriptions are stable and savable." She argues that the idea of savability and stability of inscriptions arises "socially as well as perceptually", giving the example that print media were seen as fixed more because of "early modern print circulation" than because of "any perceptual or epistemological conditions inherent to printed editions in distinction from manuscript copies." (Gitelman, 2006, 6) For Gitelman, then, as we have found, media, as well as being the seat for a representation, are material and find their place in a cultural context as a result of their materiality.

So we are left with a situation in which media are the real physical, culturally significant, evidential, tangible end of the succession of concepts we've been building up. It's in their mediatised manifestations that digital music encoding methods must assert themselves into their cultural context in order to secure any cultural credibility or longevity. The long and complex process of displacement which the music has been through before it gets into a medium would indeed be wasted otherwise.

## 3.4 Preservation and Dissemination

So finally we've displaced music into a physical and potentially durable form. Now we consider how that form can guarantee its cultural context by providing for its longevity and distribution.

First, we return to Gitelman who describes how media age, by drawing a distinction between them and works of art, on the one hand, and scientific theories on the other. She argues that old works of art "remain meaningful" whereas old scientific theories tend to become wrong, "myth or fiction". Media fall somewhere between the two in how they are affected by their temporal distance from their moment of creation. She argues that old media (she gives medieval manuscripts and computer punch cards as examples) age by becoming less able to do the work of representation. "Media are so integral to a sense of what representation itself *is*, and what counts as adequate—and thereby commodifiable—representation, that they share some of the conventional attributes of both art historical objects and scientific ones." So a musical document preserved on a musical medium is most meaningful within the cultural context in which it is created. As it moves away from this originating cultural context, according to Gitelman, it will become less meaningful and therefore less capable of being a representation of a musical work.

Similarly, Smiraglia (2001) argues that items in archives suffer from only being meaningful given an appropriate context in which to understand them.

Dodge (2002) describes how technology has affected archival practice. He argues that "too often" technology has been seen "solely in terms of its ability to solve problems—problems of access and problems of preservation." (Dodge, 2002, 21) And that "the Internet is perceived as a neutral tool for doing familiar things in a new way rather than as an entirely new environment, producing new products, new kinds of information for new uses and new users (in effect producing new cultural communities)" (Dodge, 2002, 22). He argues, instead, that archivists using technology should be aware of potential problems such as excluding groups from information, distorting the provenance of artefacts, and being more responsible for the historical narratives surrounding artefacts as a result of the much greater amount of material made available in one place and the ease with which artefacts may be juxtaposed with each other and with meta-narratives such as annotations.

These ideas bring us back to Latour and his confidence in inscriptions being able to gather, arrange, and provide manageable/managing abstractions for as many inscriptions as possible *in the same place* in order to "master" or "dominate" the intellectual field they represent. The management of the archive is summed up in his argument that, "[s]omething has to be done with the inscriptions which is similar to what the inscriptions do to the 'things', so that at the end a few elements can manipulate all the other on a vast scale. The same deflating strategy we used to show how 'things' were turned into paper, can show how paper is turned into *less* paper." (Latour, 1990, 20)

Floridi (1999), in his work on constructing a philosophy of computing, emphasises the importance of computers in being managers of information: "[t]he managerial function of computers, rather than their immediate mathematical applications, has represented the real revolution in our culture, and it is only by relating it to the history of the infosphere that the impact of ICT on our way of living and thinking can begin to be fully appreciated." (Floridi, 1999, 97)

Such managerial functions have become increasingly important in making archival material available over digital networks. The kinds of tropes which are now very common in library science discourse are those of standardisation, interoperability, and durability. As we'll see later, some key technologies employed in enabling interoperability are things such as meta-data standards, agreed upon sets of descriptive components shared by numerous digitally published archives.

Finally, we examine some attitudes of contemporary musicology to digital archives. Cook (2005, 4) argues that musicology is a "data-poor" discipline, it routinely deals with very small data sets. This has become the case, he argues, because old pre-2$^{\text{nd}}$ World War confidences in the universality of musical style and the consequent intellectual project in proving such a notion through comparative study has been abandoned on the grounds of being too nationalistic. He describes how this project was replaced by the two disciplines of ethnomusicology and structural analysis, both of which avoided comparative studies; ethnomusicology because it held that music's were rooted in cultures and comparison between cultures was illegitimate, and structural analysis because it held that pertinent properties of musical works were properties of those works only and comparison between works was illegitimate. Cook, on the other hand, argues that "*working with larger data sets will open up new areas of musicology.*" (Cook, 2005, 5)

Much of the most recent high profile applications of computers in musicology has been in archiving and dissemination over digital networks. British culture has a strong disposition for heritage and its preservation, and so any work which seeks to enable the preservation of items of significant heritage status currently attracts the attention of those involved with determining the direction of academic practice.

# Chapter 4

# Survey

Having examined some of the issues involved with making digital representations of music, we now proceed to survey a small subset of some existing and past methods in order to get an idea of the cultural practice of music encoding.

There is quite a body of prior work in analysing music encoding methods, most of which can be categorised into two groups: academic classification systems including some representative examples; and non-academic, often Web-published lists. Examples include Selfridge-Field (1997), which Selfridge-Field describes as a "tripartite categorization—codes for sound applications, codes for notational applications, and codes of analytic and/or more abstract applications." (Selfridge-Field, 1997, 28)

Wiggins et al. (1993, §2) describe two broad "orthogonal dimensions" of categorisation of encoding methods: "'expressive completeness'" and "'structural generality'". "'Expressive completeness' refers to the range of raw musical data that can be represented, and 'structural generality' refers to the range of high-level structures that can be represented and manipulated."

Huron (1992, 13) describes seven categories of music encodings:

- discrete sound functions (DSF) which is recorded sound.
- sound synthesis information (SSI) which is like an abstraction of DSF, like algorithms for, e.g., violin synthesis as well as more general synthesis tools like MUSIC V.
- performance activity information (PAI) which is gestural information including MIDI.
- common musical notation (CMN) which covers all forms of notation from common Western through paper-based visual notations of other cultures, through tablatures,

hand-signs, and Braille to sonic signifiers. He emphasises the distinction between the abstract concept of the notation representing a work and any visual manifestation of it (e.g. manuscripts and typeset notation being different, but both representing the same thing).

- visually-rendered notations (VRN) including things like SCORE, Don Byrd's SMUT, Postscript and facsimiles and image data. Huron calls the process of rendering music notation to a VRN "text setting".
- score(-based) analysis (SA) which includes Schenker graphs, Roman numerals and Forte's set forms as well as "perceptual and cognitive models".
- meta-scores (MTS) which seem to be almost like templates for scores ("class of compositions") as he describes a process of "instantiation" to produce CMN from MTS and references a few examples: "Xenakis-like tendency masks, tables of conditional probabilities associated with information-theoretic analyses, self-similar or recursive processes, transformational-generative grammars, etc."

Huron (1992) notes that the classes are "suggestive rather then exhaustive". He notes that "musical skill is in evidence in both the representations and the transformations." He describes the idea of a "broad representational network" easily "traversed" and allowing for "creative" extension including "the discovery of new goals"(Huron, 1992, 15).

## 4.1 Standardisation

One important feature of many of the practices surveyed here is that they attempt to define a representational method which they hope will find applicability beyond their own initial (and often merely exemplary) use cases. To help to achieve this, they appeal to the idea of standardisation, publishing specifications on how encodings compliant to the representation must be. We briefly consider something of the nature and practice of standardisation.

Bowker and Star (1999) propose several "dimensions" of standards: "1. A 'standard' is any set of agreed upon rules for the production of (textual or material) objects. ... 2. A standard spans more than one community of practice (or site of activity). It has temporal reach as well in that it persists over time. ... ¿3. Standards are deployed in making things work together over distance and heterogeneous metrics. ... 4. Legal bodies often enforce standards. ... 5. There is no natural law that the best shall win. ... 6. Standards have significant inertia and can be very difficult and expensive to change." (Bowker and Star, 1999, 13–14) As with their rules of classification (considered in a different context above), they argue that these rules are "idealized" (Bowker and Star, 1999, 15).

They note that standards have to be negotiated by affected parties and argue that the "history of relations" (Bowker and Star, 1999, 15) between those parties will be an important factor. Both this idea and their point 5 above are in evidence in one case of digital music encoding practice. The *MusicXML* encoding method has found some considerable popularity amongst vendors of music notation editing software packages as a result of its author's cooperation in developing software tools to allow integration of the encoding into those software packages and his enthusiastic promotion of the method to other parties involved in digital music notation tools. At the same time, however, some critics highlight important shortcomings in the encoding method's representation.

Bowker and Star (1999) also talk about the relationships between classifications and standards, arguing "[c]lassifications may or may not become standardized." (Bowker and Star, 1999, 15) So the encoding methods we examine below can be considered a kind of standardisation, a making formal of a classification which, in turn, must be made physical in an inscription. Just as the classification process displaces the music, and the inscription displaces the classification, so the standard displaces the inscribed music as yet another inscription in a Latourian cascade.

## 4.2 Making Taxonomies

The survey we carry out below is essentially a kind of taxonomy (in a loose sense). So we now briefly consider the nature of taxonomies. There are two classes of taxonomy: "scientific" and "folk" taxonomies. Scientific taxonomies make some attempt at objectiveness and universality whereas folk taxonomies take into account their own social context, they are explicitly aware that the classification in which they engage is subject to the opinions, experiences, and prejudices of the taxonomists doing the classifying. Brown et al. (1976) suggest that folk taxonomies are more concerned with language and the conventions of naming than they are to do with socially constructed classifications.

More formally, taxonomies are hierarchical. They arrange things into tree hierarchies with parent-child relationships between elements. Each subordinate *taxon* (taxonomical unit) embodies additional constraints on classification (e.g. "vehicle" is superordinate to "car", the former admits of more compliant classes than the later).

In what follows we have made a classification scheme for music encoding methods which we will examine in more detail shortly. The scheme happens to be based roughly on use-case, but many other bases would have been equally credible, such as symbolic/audio, musical genre bias, programming language/digital inscription technique, academic/commercial sector.

## 4.3 Criteria for Assessment

Here we present a list of the kinds of questions we ask of each encoding method considered:

- The levels of displacement in which it engages.
- The inscriptive and cascading (or meta-inscriptive) qualities of its representation.
- What abstractions of music does it employ?
- What are its intended objects of representation? (i.e. what music is it for?)
- Does it make provision for the musical object's (or document's) cultural context? For example, does it assume that any content to be encoded is a through-composed musical work with elements presented in a direct sequential order? Or does it make provision for encoding, for example, non-performed scale formulae, musical treatises, incomplete fragments, or documents which provide just a context for improvisation?
- How does is segment/discretize the musical information? (Does it segment it?)
- What critical apparatus has been (should be, could be) employed in transcribing the originating source (be it notation, audio, etc.) into the encoded document?
- Where does it situate: its designers? Its users?
- Who is its intended audience?
- What assumptions does it make about the skills/knowledge of its users? Does it require them to be able to understand the process of converting the musical signal

into its particular abstraction? Does it provide some sort of intuitive method for data ingestion?

- Who is/was involved in its proposing, implementation?
- What are its stated applications?
- Was it the result of a (funded?) research project?
- Is it affiliated with an institution? An archival project?
- Does it have commercial applications?
- Is it intended for dissemination? At what scale?

## 4.4 Overview of Methods

Following Cole's (1974) argument that notations reflect the interests and concerns of the cultures which develop and employ them, so too will digital music encoding methods reflect the concerns and contexts of their originating cultures. MIDI is biased towards computer performance, *MusicXML* towards notation interchange, and *MEI* towards document transcription. As a result certain comparisons between methods may be illegitimate and so we have chosen four broad categories to help alleviate any such unfair comparisons.

The practices under scrutiny in this section fall into several different categories of completeness and applicability. Some are complete *frameworks* for doing things like computational musicology, music information retrieval, music informatics. Some are completely abstract representation specifications. Some are concrete digital notations, but don't come with any particular application tool-set. Some are music programming languages.

The four categories we have chosen are: *document-centric encodings* (those which are generally notationally biased and take the musical document and its features as primary objects of representation), *framework encodings* (those which are specified as part of a tool-set for any kind of computer music processing), *bibliographic encodings* (those which deal with musical works and their various manifestations as their primary object of representation), and *performative encodings* (those which privilege the temporal aspects of music and often make provision for generative processes).

## 4.5 Document-centric Encodings

The document-centric encodings are those which deal with documentary manifestations of musical works as their primary object of representation, most often something akin to scores.

Most of the encodings in this group make some tacit, but quite important assumptions about musical works. They tend to assume that the work is a complete object which can be captured and specified through a notational representation (though some also allow for things like alternate readings or mixed sources for works). Most tend to assume a one-to-one mapping between musical work and encoded entity (or *file*, most often).

We restrict our classification of musical documents to those which are notation based, but not to any particular notational system. In fact, many of the encodings in this group employ a discretization method closely modelled on CWMN, treating the *note* as a fundamental atomic unit of music. Their structural divisions also often follow CWMN, employing entities such as *bars* (or *measures*) and sometimes things like *phrasing*. Some also define very document-centric structures like *lines* and *pages*.

As these methods are score-based, they are often involved with the process of re-inscription of existing musical texts, or *transcription*. As a result, as well as making provision for encoding differences in readings, some of the methods allow for encoding bibliographic details of sources used, and even for encoding some details of the fact that the encoded entity is a transcription (e.g. the name of a party responsible for the transcription, date of the transcription, etc.) Such provisions allow these encodings to take on an importantly cultural context-aware status; they are explicit about their participation in the chain (or "cascade") of re-inscriptions which make up the life of a musical text.

These encodings include some of the prime examples of those which privilege human readability. Some are intended to be used as input methods and so, for these, readability is clearly important. Others, however, are intended as interchange formats or to be edited using notation manipulation software and so have less need to provide for readability.

Related to this, and quite possibly because of their general textual bias, many of this class of encodings are XML-based. As a result, they've had to employ various methods of resolving XML's single tree hierarchy and music's inherent multiple hierarchies.

They find applications in score editing, publishing, as interchange file formats, archiving and dissemination, and some musicological projects.

Examples of some of the methods we class in this category include:

- *MusicXML*: an XML DTD for representing CWMN intended primarily as an interchange format between score editor applications. Conceived and maintained by Michael Good, *MusicXML* has arguably gained critical mass in its application area, partly because it is supported by the two major score editing applications.

- *MEI* (the *Music Encoding Initiative*): an XML DTD for representing musical documents more generally. Conceived and maintained by Perry Roland in relation to Indiana University's digital music library research project, *Variations*. (Described in detail below.)

- *CMME* (*Computerized Mensural Music Editing*): an XML DTD for representing mensural music notation (a notation commonly used in Europe for vocal polyphony from about the thirteenth to the sixteenth centuries). It was conceived as a student project by Theodor Dumitrescu, later developed under post doctoral research and is currently housed at the University of Utrecht and funded by the Netherlands Organisation for Scientific Research[1]. The current project under which the encoding method is developed is attempting to produce scholarly editions of works of medieval and renaissance vocal polyphony to be published online. The project also includes the development of software tools for editing and viewing CMME data. The encoding method is remarkable in its provision of semantics particular to mensural notation conventions and its provision for encoding differences in source documents.

- *DARMS* (*Digital Alternate Representation of Musical Scores*): an early attempt at designing a text-based encoding method for CWMN. It is oriented towards graphic elements of CWMN and prioritises keyboard input. (Described in detail below.)

- *NIFF* (*Notation Interchange File Format*): is an attempt dating from around 1994 to produce a standardised representation method for CWMN to allow musical notation to be shared between the various different notation editing software packages then available, and particularly the upcoming optical music recognition programs. It privileged page layout concerns, but also made provision for encoding performance information. It followed the Resource Interchange File Format (RIFF) rules proposed

[1] http://www.nwo.nl/ Accessed 10 May 2009

by Microsoft and IBM in 1991 in which file formats must include meta-data such as chunk sizes within the files in an attempt to make provision for the enduring usability of the encoded data. NIFF was quite flexible in what it could represent as a result. Unlike the other representations discussed in this section, NIFF was a binary format (that is, it was not text-based and not human-readable).

- *TabCode*: (Crawford, 1991) a plain text encoding for representing lute tablatures. Like *CMME*, *TabCode* is importantly application specific in that it makes provision for encoding a specific kind of musical data. Unlike CWMN, tablatures are not diastemic and, as representing diastemical notations within the constraints of computer files is difficult, methods such as *TabCode* have an immediate advantage. Tablatures encode instrument instructions rather than (as in CWMN) pitch. In the case of lute tablature, those instructions are which string/fret positions the fingers should be placed on. *TabCode* (like the written tablature) discretises the notation into time chunks, each chunk representing a simultaneous arrangement of fingers. The chunks in *TabCode* are called `TabWord`s and are separated by white-space (space, carriage return, tab space, etc.). Each `TabWord` consists of a string of ASCII characters which denote which notes should be played and optionally a duration value for the `TabWord` expressed using a mnemonic for American rhythm names. The *TabCode* representation has been applied in the ECOLM (Electronic Corpus of Lute Music) project which aims to make available online to scholars and performers editions of lute repertoire encoded using *TabCode*.

- *GUIDO*: a plain text encoding for CWMN. GUIDO explicitly privileges simplicity of representation of the most common elements of CWMN and values human readability highly.

- *Lilypond*: a music type setting system that uses a TeX-like input method. (Described in detail below.)

### 4.5.1 The Music Encoding Initiative

In November 1987, the Association for Computers in the Humanities[2] and the National Endowment for the Humanities[3] sponsored a meeting held at Vassar College to address

[2] http://www.ach.org/ Accessed 10 May 2009
[3] http://www.neh.gov/ Accessed 10 May 2009

problems then current in producing, manipulating, and maintaining digital research data for humanities scholarship. The primary issues under discussion were longevity and re-usability of textual data, issues made difficult at the time by the proliferation of ad hoc or proprietary encoding methods and tool sets. The outcome was the establishing of the Text Encoding Initiative (TEI), a set of guidelines for the digital transcription of textual documents. In 2000, a TEI Consortium was incorporated which took responsibility for overseeing the continuing development of these guidelines to maintain their currency and applicability and for promoting them to humanities scholars.

The TEI consists of a set of schema organised into modules which can be combined and extended according to the needs of the user. The schema (expressed using any of the known methods for encoding XML schema, as described above) contain definitions for the classes of elements which may appear in any document conformant to the schema, for the attributes those elements may have, and for the relationships instances of those element classes may have with each other. The schema modules defined in the standard TEI cover a wide range of practices in textual scholarship including: describing manuscripts, building corpora, analysis and interpretation, dealing with dramatic and verse texts, encoding unusual glyphs and characters, describing transcriptions of primary sources, and embedding textual criticism. As a result of the ease of what Latour calls "superimposition" of digital inscriptions made possible with markup methods such as XML (see above), the TEI is able to provide scholars with the tools to encode all these properties of texts orthogonally (actually in opposition to Latour's "flatness"), something which is much more difficult and cumbersome in print publications.

Perry Roland's Music Encoding Initiative (MEI) attempts to do for digital music transcription what the TEI has done for digital text transcription. Roland describes as an important property of the TEI the fact that it is not "'the human expression encoding initiative'" (Roland, 2002, §2). Although its source domain is text and text can include numerous aspects such as logical, visual, aural and analytical, it is deliberately limited to "those expressions which take a *written* form." Similarly, Roland argues, his MEI only attempts to encode music as a written phenomenon, it only deals with the notational inscriptions of music and, importantly, excludes any provision for dealing with how music is created.

Roland describes various design principles behind the MEI which include some related to how it represents music such as being hierarchical and declarative, but also some related to how it should fit into a cultural practice. For instance, Roland argues that the encoding should be make clear reference to the fact that it represents the end result of a process of interpretation and provide mechanisms for encoding interpretative choices made. Similarly, he argues that the encoding should embody what he sees as the opposing properties of being “formal” and “flexible”. It should be formal in that it should be possible to “prove [its] correctness” (Roland, 2002, §3) without reference to any knowledge outside of the specification, and yet flexible enough to allow those who make documents using it to encode variations and to ignore features of the source material which are not of interest.

In implementation, the MEI, like the TEI, consists of modularised XML schema specifying groups of features of musical documents such as: meta-data, corpora, physical properties, segmentation, temporal (horizontal) and concurrent (vertical) relationships, variant readings, and control events. Documents may make use of any of these modules and of the element classes they contain.

MEI separates the meta-data for a musical document from the musical data by using two high-level container elements: `<meihead>` for the meta-data and `<work>` for the musical data. It not only allows for a rich set of meta-data elements, but also for those elements to indicate their equivalence with elements from any other meta-data encoding standard (such as Dublin Core[4]). Thus, MEI data may be integrated into other meta-data catalogues. Its meta-data elements include things such as details on availability, publication, edition, and access rights of the source document. They also include some more reflexive properties such as details on why the music has been encoded, its status within an editorial project, and date and personnel information regarding document encoding. It allows for detailed information on source documents such as physical description, provenance, the *hand* in which it is written, and any restoration or exhibition details. It allows for multiple source documents to be described and also includes a mechanism by which some structural elements used in encoding the musical data may refer to source documents. Borrowing from TEI, it adopts the term “statement of responsibility” for concepts such as composer, author, and editor. All these demonstrate a strong understanding of the part an electronic edition plays in the history and cultural construction of a musical document.

MEI is very open about what constitutes a musical work and allows single tran-

[4]http://dublincore.org/ Accessed 10 May 2009

scriptions to include anything from fragments to whole corpora though its `<group>` and `<meicorpus>` elements. On the other hand, it is very clear about the musical document's status as a text. It borrows the `<front>` and `<back>` elements from the TEI which can be used to include front and back matter text in the encoded document. Musical data itself is encoded inside a `<music>` element.

MEI takes two broad approaches to discretization of music into larger, higher level structures. The first is its `<mdiv>` element which can be used to encode generic top level divisions such as movements of symphonies or acts of operas. As part of this discretization scheme it uses a `<section>` element which is agnostic of any particular semantics but functions mainly as a delimiter for the scope of application of properties such as meter and key as well as arbitrary user-defined properties. Notice how, despite the ability to nest `<section>` elements, these provisions are essentially linear in nature, one element follows another. MEI's other broad discretization mechanism attempts to overcome this linearity by flattening one of the dimensions of music, concurrent parts. It uses `<score>` and `<parts>` elements inside `<mdiv>`s to allow musical data be to encoded as either a temporally ordered sequence of concurrent instrumental events (a `<score>`) or as individual instrumental parts (using `<parts>` and `<part>`).

Middle level discretization is handled by elements inside the `<section>` elements and, like the score/part division described above, requires a flattening of the two-dimensional source document into a sequence of linear, serial elements either by *staff* or by *measure*[5]. That is, `<measure>` elements containing events data may appear in sequence inside `<staff>` elements, or the encoder may use a sequence of `<measure>` elements which each contain `<staff>` elements. Although event elements (see below) may appear directly beneath `<measure>`s or `<staff>`s, the MEI also provides `<layer>` elements which allow concurrent *voices* to be encoded.

The finest level of discretization defined in the MEI is what Roland calls "events", "the typical, time-based, discrete atoms of musical data, such as notes, chords, rests, etc." They are encoded as empty elements rather than as character symbols because, Roland argues, to use symbols would "place too much emphasis on presentational qualities" (Roland, 2006, 6). The properties of these events are encoded using XML attributes. `<note>` events, for example, may have attributes such as `pname` (the letter name of the pitch class), `oct` (a

[5]American musical terminology is adopted in the MEI.

number representing the octave of pitch to make the pitch class absolute), `dur` (a number representing the duration of the event expressed as the inverse of its proportion of a semibreve) and various presentational attributes such as stem direction. Events may also have a unique identifier attribute (`id`) so that they can be referred to unambiguously elsewhere in the document. For example, some small scale structural features such as *beaming* of durationally equivalent notes or phrase markings may be encoded as displaced elements (e.g. `<phrase>`) which refer to the notes they encapsulate using attribute values (e.g. `start`, `end`) containing these identifiers.

Two of the XML idioms we introduced above are at work here. First is the use of empty elements with attributes to inscribe atomic entities, as seen in the `<note>` element. As we discussed, there is no pre-existing content (characters) in the document which the `<note>` element can be applied to. Rather, the element itself is the content. There are also several examples of encoding conflicting hierarchies using a type of stand-off markup. The concurrent `<staff>`s and `<measure>`s constitute a conflicting hierarchy: staff and measure boundaries do not align and so it's not possible to nest one inside the other in markup consistently. Similarly, phrase markings and note beaming often cross hierarchies such as measure divisions and so, in MEI, are encoded using displaced elements which refer back to the event objects they apply to.

One of the most important features MEI borrows from the TEI is being able to encode differing readings of a source musical text in the digital edition. Like the TEI, it uses an `<app>` (apparatus) element to enclose two or more `<rdg>` (reading) elements. Each of the `<rdg>`s contains musical data for the same portion and may be linked to a source document listed in the header. MEI also provides the `<ossia>` element for encoding alternatives which are actually in the source document (rather than being variant readings by the encoder). These provisions, along with the `<front>` and `<back>` matter elements, and the meta-data's strong emphasis on source documents, are examples of the narrative-like or document-centric class of XML markup paradigms discussed above. However, MEI also makes considerable use of more record-like or data-centric idioms in its handling of event data as sequential, self-similar objects.

There are several examples of displacement at work in MEI encoded documents. One is that, by encoding conventional notation, it encapsulates the displacing that notation already does; breaking the sonic continuum into discrete pitch events and assigning symbols to them. It also displaces the source document as an entity by being a transcriptive practice.

MEI has been employed in various transcription projects including the *Online Chopin Variorum Edition* (OCVE) run by John Rink at Royal Holloway, University of London which attempts to make available online tools for comparing sources of some of the piano works of Chopin. MEI is used to store symbolic representations of the musical information in the sources. Those sources are also made available in graphical representations and the alignment between the logical notational data and its graphical notational equivalent is explicitly encoded, thus allowing the retrieval of manuscript and printed sources from abstract notational queries. The project also aims to integrate some interesting extra-musical information into its available sources, including scholarly commentaries and annotations. MEI's flexibility in extending its representational range makes it a suitable method for encoding such information.

### 4.5.2 DARMS

The Digital Alternate Representation of Musical Scores (DARMS) is one of the oldest digital musical encoding methods. It dates back to 1963 in an originating proposal by Stephan Bauer-Mengelberg for a method of typing musical notation with a computer keyboard and was still in use up to the mid 1990s. The definitive reference specification for DARMS (often called *canonical DARMS*) is Erickson (1976). Several derivatives or dialects were developed, mainly in connection with music publishing houses or specialist applications such as for lute tablatures and mensural notation.

DARMS is very much oriented towards the representation of the graphical elements of CWMN and consequently most of its applications are in the production of printed scores. As another consequence, though, it has quite a distinctive way of dealing with the common musical concepts.

DARMS does not represent pitch so much as it represents the vertical placement of a note-head on a staff. This is encoded by assigning each line and space of the traditional five-line musical staff a number: 1, 3, 5, 7, and 9 for the lines and 2, 4, 6, and 8 for the spaces[6]. The numbers then extend in both directions for note-heads above and below the

[6]In fact, this is the commonly used abbreviation for staff line numbering. In canonical DARMS, each staff for multi-staved instruments (such as piano) is assigned a different set of fifty numbers with the top-most (for example right hand on piano) taking 21 for its bottom staff line, the second staff down (for example left hand on piano) taking 71 for its bottom staff line, etc. That way, each line or space for the

staff (i.e. for higher and lower pitches).

Reflecting its predisposition for notation, DARMS data is very much centred around the concept of the note. Notes may have up to nine common properties: 1) vertical space code (pitch, as we've seen); 2) pitch alteration code (or chroma); 3) note-head code (also used for rests); 4) duration code; 5) horizontal space code; 6) tie code; 7) stem code; 8) beam code; and 9) performance codes (such as slurs, articulation marks, dynamics, etc.).

In keeping with its intended application for typing music, the codes used for these properties are extremely terse, often consisting of just one or a few characters, and as many of them are often not needed, notes may be represented very succinctly. Furthermore, DARMS makes provision for even greater brevity by allowing notes to leave out any properties which are identical to their predecessor's. For example, if four successive eighth notes (quavers) are to be encoded, only the first need include the code for eighth note (`E`) and the following three can be represented using just their vertical placement values (pitches). The note and its necessary properties are then inscribed as a group of these characters with spaces used to separate character clusters representing different notes.

DARMS caters for several other note properties. Pitch alteration or chroma is encoded with `#` for sharp, `##` for double sharp, `-` for flat, `--` for double flat, and `*` for natural. Alternative note-heads are encoded as `N` followed by a number including `0` for no note-head, `4` for square note-heads, and `R` for rests. Durations are represented by a mnemonic letter: `W` for whole (semibreve), `H` for half (minim), `Q` for quarter (crotchet), etc. Tied notes are represented by the code `J` (for join) and, most simply, tie one note with the next note to have a `J` in its inscription. Multiple (and overlapping) ties can be represented by appending the `J` with numbers: starting tied notes are assigned odd numbers and their corresponding ending notes are assigned values incremented by one. Slurs are represented in a very similar manner except using `L` rather than `J`. Stem direction may be specified with either `U` for up or `D` for down. Rhythmic beaming is represented using parentheses: an open parenthesis is included in the inscription of the note which starts the beamed group and a close parenthesis in the inscription of the note which ends it, e.g. `3E( E)`. Various types of articulation may be encoded using symbols such as `>` for accents, `'` for staccato, or `_` for tenuto. Dynamic markings are represented using the code `V` followed by the Latin letters used in CWMN notation to notate dynamics, `F` for *forte*, `MP` for *mezzo*

instrument has a unique number. Also in canonical DARMS, single-staff instruments would take 21 as the number of the bottom staff line.

*piano*, etc.

DARMS employs *global specifiers* for score properties such as time and key signatures. Global specifiers are inscribed preceding the portion of notes to which they apply. Examples include `!G`, `!F`, and `!C` for the G, F, and C clefs respectively, `!K` for key signatures where the number of flats or sharps is represented, for example, as `!K3-` or `!K4#`, and `!M` for time signatures, for example, `!M4:4` or `!M12:8`.

It encodes no more musical structure than CWMN, allowing the representation of bar (measure) divisions using `/` and repeat marks using `/:`, `:/` and `:/:`. It does, however, allow for the encoding of repeated patterns using the codes `!X`*n* and `$X`. Anything between the two delimiting codes is repeated $n$ number of times. This construct, however, does not allow parameterisation of the enclosed DARMS code so, for example, ascending or descending musical sequences can't be encoded this way.

Like many of the text-based encoding methods, DARMS avoids dimensionality in its music representation in several ways. As we've seen, it doesn't follow CWMN's method of representing pitch spatially, but it does represent time spatially, at least by encoding notes sequentially (if not proportionally). However, this breaks down at representing concurrent notes. To encode chords it simply separates notes with commas rather than spaces. To represent multiple voices on one staff it provides a mechanism called *linear decomposition*. Notes which are to be linearly decomposed are delimited by the codes `!&` and `$&`. A series of DARMS encoded notes are then inscribed between these marks and each occurance of a `&` in these notes indicates that the following notes should be treated as a new voice and occur simultaneously with the notes from the other voices.

DARMS borrows some of the displacing power of CWMN. Just as CWMN displaces through discretization the musical continuum, so DARMS, by re-representing CWMN's *notes*, borrows this same displacement process. However, as we have seen, DARMS does not attempt to employ the same spatial metaphors in its representation that CWMN does.

An important implication of many of DARMS's techniques is that they are often very unsuitable to automated analytic applications. For example, if pitches and durations were represented using a numerical system which correlated to some numerically manipulable properties of sound or time, then music analytic operations could be modelled mathematically. (The credibility of any such models would be another matter.) As the codes are chiefly mnemonic and privilege a lack of redundancy, however, even statistical operations (such as simple frequency analysis) are non-trivial.

DARMS's emphasis on being a digitisation of CWMN, however, found applications in the production of printed music. J. Stephen Dydo's *Note-Processor* was available at version 2.0 in 1989 (see Skinner, 1990, 660) and used DARMS as its input method to produce printed sheet music. Similarly, *A-R Editions* has used a dialect of DARMS devised by Thomas Hall for music publishing since 1981 (see Hall, 1997, 193). Dydo and Hall worked at Columbia and Princeton Universities which, along with Bell Labs, were some of the pioneer institutions in the application of computers to music in the 1950s. The *COMUS* music printing software was developed by John Dunn from the 1970s through to the 1990s and, again, uses a dialect of DARMS[7]. Despite the shortcomings we covered above, there have been various music analytic applications of DARMS. Alexander Brinkman's text on doing music analysis with the *Pascal* programming language uses DARMS as a representation method (Brinkman, 1990, 137–154). Erickson himself employed DARMS in analysing rhythmic features of twelfth century chant.

### 4.5.3 Lilypond

The Lilypond software is a music typesetter which aims at emulating what its designers refer to as the "beauty" of hand-made music engraving. They oppose this style to the more common precise and "mechanical" engraving methods they argue are common in most mainstream score-editing packages.

As a result of this pre-occupation with the layout of notational elements, Lilypond has some very particular concerns regarding the requirements of a music encoding method. They argue that, "music notation is really different from music itself. Notation is an intricate symbolic diagramming language for visualizing an often much simpler musical concept. Hence, software should reflect that separation." In order to make provision for their argument that "content and form of a score are separate", they implement a representation scheme which divides musical score information into three parts: "typography: *where* to put symbols; notation: *what* symbols to produce; representation: how to *encode* music." Note, however, that these components are implemented procedurally rather than declaratively.

Lilypond, as software, embodies are great deal of procedural knowledge. Its designers

[7] http://comusprint.co.uk/COMUSdesc.htm Accessed 10 May 2009

have organised it into separate modules (or "plug-ins") for handling different aspects of the engraving process. For example, note-heads, staves, clefs, stems, accidentals, bar lines, key signatures. Each of these modules specialises in how to inscribe one small aspect of common musical notation and many of them have rules for interacting with each other. For example the clef, staff and note-head modules must work together to discern the correct vertical placement of a pitch, and the accidental and clef modules must work together to calculate when a chroma specification sign is necessary.

The Lilypond engraving system also makes provision for some of the conflicting hierarchies present in musical information. For example, the functioning of its basic modules becomes much more involved when the music being engraved is polyphonic. In this case it uses two *context* modules: *voice context* and *staff context*. The voice context deals with properties which apply only to one line or voice within a staff (such as stems, slurs, and beams) while the staff context deals with properties which are shared between voices (such as accidentals and bar lines).

Lilypond also employs a text-based representation for typing in musical data to be typeset into CWMN. This encoding method relies on the phonetic (as opposed to diastemic) affordances of plain text encodings for musical information, using letters to represent pitch class and numbers to represent rhythmic values. It takes advantage of various visual cues allowed by basic ASCII characters for delimiting groups in order to represent grouping concepts in music. For example, it uses matching pairs of brackets preceded by *escaped* descriptor strings to delimit structures such as staffs: `\staff { ... }`[8].

## 4.6 Framework Encodings

The framework encodings are those which are specified as part of a tool set for working with musical data in computers, often music analysis or composition.

Like the document-centric encodings described above, they often deal with segmented music and take their segmentation scheme from CWMN. However, unlike the document-centric encodings, they often make explicit provision for dealing with the segmented data diachronically (e.g. encoding *events* against *timelines*).

They usually don't make any provision for encoding or preserving details of the cultural context of the music. Their conception of music is usually much more as *data* than,

[8] This notational style is in fact borrowed from the TeX text typesetting system.

particularly, the document-centric encodings. For example, they don't preserve details of any documents from which the music may have been transcribed, but rather take some abstraction of musical information (often CWMN) and require that all musical data is framed in terms of that abstraction.

They are often designed as part of a tool set for computational music processing and so may be quite restricted by that tool set. For example, they may encode musical elements in terms of data structures used in the programs which make up the tool set, or be required to organise their files in a way which aides the efficiency of those programs, or rely on some data such as timing synchronisation which only exists while programs of the tool set are running.

Often these encodings (and their associated tool sets) are targeted at both computer scientists and musicologists but only appeal to computer scientists and those few musicologists who are interested enough to climb the steep learning curves they present. Similarly, Cook (2005, 6) argues that the current state of tools for computational musicology is a disbar to uptake. He argues that the number of different tools is "daunting" and that musicologists would probably like one all-purpose tool, but that this will not happen because there is no infrastructure to develop and distribute one (e.g., he says, there is no "Microsoft"). He praises "current developments in linking and annotating different representations of music". Similarly, he argues that "unified interfaces" to tools and representations (he refers to OMRAS2) are a good step.

Some examples of framework encodings include:

- *Kern*: the encoding method used in David Huron's *Humdrum* toolkit for computer assisted approaches to musical research. Huron (1992, 36–37) outlines some concerns he feels are important in designing computer music representations. In his reasoning, he values things like disk space and takes care to utilise the limited set of "signifiers" well. His solution is somewhat like stand-off markup: instead of using up all the available characters in specifying what their interpretations should be for musical representation, you merely design protocols for how generic alphanumeric representations should interact with each other. *Humdrum* is Huron's protocol and is essentially that characters appearing in the same line in a text file represent objects which are temporally aligned, and the semantics of any given character is dependent on the *column* in which it appears, allowing a character to be re-used in a different column (context). As a result, the Humdrum representation is a plain-text file arranged in columns, with time extending downwards, and employing phonetic representations of pitch and rhythm. However, Huron only provides the syntax for the more common types of musical information which users may wish to encode (such as pitch, rhythm, harmony), relying on the underlying *protocol* of Humdrum (horizontal temporal alignment) to allow users to invent their own representational methods for any other type of information they may require. Huron then relies on tools for processing textual information provided by the computer operating system (such as pattern matching programs) for working with the musical information.

- *MUSITECH*: a framework for encoding and processing musical information designed for use in the music informatics research group at City University, London. (Described in detail below).

- *AMuSE*: a set of generic operations or functions[9]which describe the kinds of processes that may be applied to musical information, for example, manipulating time and pitch intervals. The operations are specified in an abstract manner but may also be implemented concretely to work with different kinds of musical representation such as digital audio or MIDI instructions. Therefore, AMuSE also includes a database of music to which these generic operations can be applied in order to implement algorithms which attempt to answer interesting research questions, such as inferring suitable harmonisations from a given melody. The framework is developed and used in the Intelligent Sound and Music Systems research group at Goldsmiths, University of London.

### 4.6.1 MUSITECH

Gieseking and Weyde (2002) describe the aim of the MUSITECH (Music and Sound Objects in Information Technology) framework as, "to conceptualize, develop, and evaluate musical interaction and navigation in a computing environment." (Gieseking and Weyde, 2002, 2) They argue that "meaningful interaction and navigation" requires the "integrat[ion] of different levels of musical information." Much of their statement of requirements of such a framework is taken up with emphasising the importance of "structures

[9]The functions are implemented in the Common Lisp programming language, but could be implemented in any language.

and relations" and allowing them to be arbitrarily defined and manipulated. They also make reference to common tropes of "extensibility", openness of data formats to ensure longevity, platform independence, object oriented implementation, and scalable in installation (available as a local application or distributed over a network).

To deal with the temporal components of musical information, Gieseking and Weyde (2002, 3) define a `TimedEvent` interface for temporal objects which has a `TimeStamp` property. They argue that this part of the model is general enough to represent music that doesn't fit into the Western note-based model. Each encoded work has a `TimeLine` object which contains references to all the `TimedEvent` objects, an arrangement which, they argue, enables "efficient temporal synchronization". The `TimeLine` object uses real time, while metrical time is encoded in "score notes". They argue that this relationship between real time and metrical time is optimal because the mapping between the two is non-linear, "complex, usually not known, and as a concept questionable."

To deal with arbitrary structure, Gieseking and Weyde (2002, 3) define collections of elements as "pools". "There are different pools for different basic object types.", including a `Music` object and a `NotePool` object, of which just one is used per "piece". These pools are importantly different from the common note-list structuring method adopted in many other encoding methods. The structural relationships between the components are off-set from the components themselves, a note doesn't have its sequential relationship to other notes as an internal property, that relationship is encoded elsewhere.

Adopting common Western musical terminology, Gieseking and Weyde (2002, 3) use the term "note" to describe both sound and written musical components. They define a `Note` object and `ScoreNote` and `PerformanceNote` objects. "`Note` objects are composed of a `ScoreNote` and one or more `PerformanceNotes`". In the framework, when one class of note information is not available, it can be generated algorithmically from the other (they mention processes like "beat tracking" and "quantization").

Gieseking and Weyde (2002, 3) argue that MIDI-like structures using keyboard-based pitch data are useful for performance applications and attempt to justify this by arguing that often only MIDI data is available. Borrowing from MIDI, they represent performance "time parameters" in milliseconds. However, they argue that "enharmonic pitch information in conjunction with meter-oriented onsets and durations are essential parts in terms of music notation and score based analysis", and so, also taking into account other notational details missing from MIDI like slurs and ties, argue that a semantically richer model is necessary.

Their representation can also include non-event elements. They give the example of `AudioObject`s which implement `TimedObject` so that their playback can be synchronized (`TimedObject` has a `TimeStamp` property). They represent lyrics, scanned score pages and video similarly as temporally synchronized objects.

Gieseking and Weyde (2002, 4) allow for the encoding of "extra-musical" information and provide "pre-defined meta-information objects like composer's name and composition date." They also provide a `MetaInfo` object which uses MIME (multipurpose internet mail extensions, a standardised method of describing the nature of digital content) to specify the type of its content.

In keeping with their emphasis on the value of arbitrary relationships in music encoding described above, they provide a generalised `Container` object. They argue that it's important not to be too prescriptive about the classes of containers available and so the meaning of a `Container` is only reified by defining appropriate properties and methods for the class of structure or relation it happens to be representing. In order to allow arbitrary depths of containment relationships, their containers can contain other containers. They argue that in this way Schenkerian analysis, paradigmatic analysis and electroacoustic music ("by referring to audio-objects") can be encoded.

Inscriptively then, MUSITECH has some interesting properties. First, it makes the graphic and sonic inscriptions of sounds conceptually equal and both subordinate to an abstract concept Gieseking and Weyde call "Note". It then doesn't admit of any concrete score-concept, but rather relies on arbitrary (or, in their terms, "flexible") relationships between inscriptions to define structures.

MUSITECH is implemented in the Java programming language and applied in internal research projects by staff and students at City University, London.

## 4.7 Bibliographic Encodings

The bibliographic encodings are those which deal with musical works and their various expressions and manifestations as atomic entities. Instead of modelling the fine details of musical events or notation-derived information, this class of encodings tend to model concepts such as musical works, publications, performance histories, movements (parts of works). As a result, they often place the cultural context of the works they model in a prominent position and privilege the capturing of details such as provenance of sources.

They often embody strong conceptions of the musical media they describe (manuscripts, printed editions, vinyl recordings, etc.) and the inscriptive practices involved in the creation of those media. However, they tend to stop before the level of detail of characters and marks.

They are often aimed at librarians, cataloguers and archivists and have arguably been the most successful in application. Increasingly, online library catalogues employ the FRBR principles (described above), and several quite large scale projects have produced a quantity of bibliographic musical data which has achieved a critical mass not yet known in the more detailed music data sets employing event- or notation-level music representations.

Some examples of bibliographic encodings include:

- *Plaine and Easie Code*: a text-based representation for describing basic bibliographic details of musical works including incipits as well as title, composer, date, etc. (Described in detail below)
- *Music Ontology*: an XML-based representation modelled on the FRBR for describing musical works and their various manifestations and production processes and intended to integrate into arbitrary online data-sets as a result of utilising standardised representational techniques. (Described in detail below)
- *MPEG-21*: an attempt to provide a standardised method of representing *digital items* (or *content*) and the ways that *users* (which may be parties of one individual or more) may interact with such items. Such interactions include being able to access, exchange, create, aggregate, publish, regulate digital items as well as other forms of manipulation. MPEG-21 provides a scheme for describing digital representations of works called the Digital Item Declaration (DID). This scheme includes a set of abstract terms for describing works and a description of the semantics of those terms. All these components are encoded using XML. The terms MPEG-21 provides include: *container* which acts as a grouping mechanism (an important contributor to its status as a bibliographic encoding in this study), *item* which is a digital manifestation of a work and a container for sub-items and the other components which DID constructs works out of, *resource* which is some digital (or possibly

physical) content and which must be uniquely addressable, *fragment* which is a way of identifying a part of a resource, and *descriptor* which allows information to be associated with an element. MPEG-21 is intended to be sufficiently general to be applicable to any form of content, though digitally encoded music will be a prime candidate for description via MPEG-21.

### 4.7.1 Plaine and Easie Code

The Plaine and Easie Code (P&EC) was initially used to encode musical incipits as part of bibliographic systems and exists in at least three forms: Plaine and Easie Code, Simplified Plaine and Easie Code (a later, "lite" version, described in Brook, 1969), and a slightly adapted version used by RISM (detailed below). In fact, its bibliographic capabilities are more limited compared to the other encoding methods considered in this section.

Inscriptively, it aims to make use of "ordinary typewriter characters" (see the title of Brooke and Gould, 1964). "It should be as closely related mnemonically to musical notation as possible, so that it appears natural and right, avoiding arbitrary symbols. ... It must be easily recognizable as music from the symbols alone and immediately retranslatable, without loss, into conventional notation." (Brook and Gould, 1964, 143) And in terms of the inscriptive process, Brook states: "[i]t should require only a single line of typewriter characters without the need for backspacing or for a second pass over the line." (Brook and Gould, 1964, 143) Interestingly, then, this encoding method was originally conceived for inscription using typewriters. However, Brook also required that it should "be so devised as to be readily transferable to electronic data-processing equipment for key transposition, fact-finding, tabulating, and other research."

Brook and Gould (1964) also make further reference to intended use and limitations. It should "be applicable to all Western music from Gregorian chant to serial music." (Brook and Gould, 1964, 143) It was proposed as a standard for use in "library card catalogues, providing quick and precise identification of a musical work by its coded incipit" (Brook and Gould, 1964, 142) and also in "research projects, assisting in search, fact-finding, organization, tabulation, identification of anonymous works, etc."

Howard (1997, 363) highlights some important limitations of P&EC's design: it's only useful for "melody in conventional staff notation" and therefore encoding anything else requires an intermediary transcription process ("depend[ent] on a critical apparatus specific to a particular application.") He also notes that it cannot handle harmony or counterpoint because of its inherent "linearity" and argues (after the authors, in fact) that it's most suitable for "musical repertories from the seventeenth through the early nineteenth centuries, or tune repertories in general."

P&EC inherits all the displacement characteristics of CWMN which it attempts to emulate within specific inscriptional constraints. It is also displaced from the musical surface or event by being a bibliographic encoding method—it doesn't attempt to represent the music in any sort of completeness, it merely attempts to be an index to or distinguisher of musical works for the purpose of being able to organise musical works. In this, it is very much along the lines of Latour's metainscriptions, cascades of summarising and simplifying inscriptions which help users to master large knowledge domains.

P&EC has been adopted by the Répertoire Internationale des Sources Musicales (RISM, founded in Paris in 1952)[10] for encoding the incipits used in the bibliographic project *RISM Series A/II*, an index to (at the time of writing) 624,952[11] music manuscripts held in over 750 libraries internationally. This database is still under active maintenance with twice-yearly submissions of around 10,000 records.

The latest version of the standard is maintained by the International Association of Music Libraries (IAML) [12] and RISM. RISM is a distributed project with a central office in Frankfurt, Germany, (the RISM Zentralredaktion), and numerous other offices worldwide including a US office at Harvard University and a UK office at Royal Holloway, University of London.

Consequently, it is necessary for RISM to manage its encoded data carefully. Howard (1997, 370) gives the example that the Zentralredaktion maintains two encodings: one "'external'" which is their P&EC-based encoding and one "'internal' encoding (or 'metacode')" which is designed to simplify the sorting, proofreading and printing of P&EC, made difficult in P&EC because its symbols do not lend themselves to simple lexicographic sorting and rely on "context-sensitive" syntax (for example, rhythmic grouping and pitch chroma being dependent on the content of the metadata fields).

Brook (1969, 292) describes an early revision of the code called *Simplified P&EC*. This version separates the data into two parts: the metadata part ("preliminary information")

[10]http://rism.ub.uni-frankfurt.de/ Accessed 10 May 2009
[11]http://www.nisc.com/factsheets/qrism.asp Accessed 10 May 2009
[12]http://www.iaml.info/ Accessed 10 May 2009

and the notation data part ("The Notes"). The metadata includes "name of the piece, the number of the movement, instrument or voice, clef, tempo, key signature, and meter" He stresses that the order they are encoded in copies the order they would normally be presented in conventional notation. He gives the cello part of 3rd movement of Beethoven 5 as an example: "`III, cello, F-4 clef (All, bBEAm, 3/4)`".

The RISM implementation has altered this metadata line into metadata *records* following the convention of various long-standing bibliographic metadata encoding methods. The version as at 1978 (described in Howard (1997, 368)) uses the records numbered `112A ... 112E` (and `112F` for the incipit). `112A` is the part number, `112B` is the instrument or part name, `112C` is the clef, `112D` is the key signature, and `112E` is the time signature.

However, the version as at 2004 described on IAML's website[13] gives these six records as numbers `820 ... 826`. This is indicative of altered inscriptional concerns: Brook was concerned with how to use a typewriter to encode this information while RISM and IAML are concerned with fitting the code into machine readable records and conforming to existing bibliographic encoding standards.

### 4.7.2 Music Ontology

The Music Ontology project aims to provide a subject-specific implementation of the principles laid out in the Functional Requirements for Bibliographic Records (FRBR, introduced above) delivered in a machine-readable form over digital networks, specifically a high-level digital network called the Semantic Web. It relies on various base technologies and assumptions about the nature of music.

One of those technologies is the W3C-recommended Resource Description Framework (RDF) which it uses to encode data and which consists of simple 3-tuples (or "triples"), the semantics of whose members are: object, predicate, subject. The predicate describes the relationship between the object and the subject. The object and predicate must be expressed as Uniform Resource Identifiers (URI, RFC 2396[14]) and the subject may be expressed as a URI or as a plain text string.

This use of resource identifiers to point to content (rather than encoding it as simple text within the triple) is the key to network delivery of the Music Ontology (and the Semantic Web in general). The hope is that any expressions which make reference to

[13] http://www.iaml.info/activities/projects/plain_and_easy_code Accessed 10 May 2009
[14] http://www.ietf.org/rfc/rfc2396.txt Accessed 10 May 2009

the same resources will use the same URI to describe them. For the object portion of a triple this idea is quite intuitive and straightforward. Where proponents of the Semantic Web argue that this model is most powerful is in the enforcement of resource identifiers for predicates—all expressions which model a particular class of relationship do so using *exactly the same* relationship-class identifying resource.

Proponents argue that encoding knowledge in this manner is beneficial because of its high degree of machine readability. Mechanical interpreters of this information ("agents") can literally follow the links in the triples as a way of conceptually following the relationships they represent and, as a consequence, build up their own internal representation of a subset of the knowledge available. Agents effectively gain an understanding of the knowledge.

The second level of semantics employed by the Semantic Web is in defining *ontologies*. Using the same representational technique as for the data, RDF, practitioners define *classes* and *relationships* which describe concepts within a knowledge domain. Just as those who define DTDs (schemata) for XML dialects (such as MusicXML and MEI described above) seek to enable interoperability of knowledge within a domain by employing standardised syntactical methods to constrain the semantics of a representational schema, so designers of ontologies attempt to constrain (or standardise) what concepts are available to those building up bases of knowledge. Here, proponents argue that the formal defining of abstract, high-level classes allows the standardisation of concepts across disparate knowledge domains, and eventually of one, all-encompassing knowledge base (the Semantic Web).

The Music Ontology, then, consists of definitions of numerous classes of objects and relationships, each with unique network-delivered resource identifiers, which represent concepts that its designers believe are pertinent to describing music, particularly the production processes of recorded music. They are keen to stress what they call its "democratic" nature. Anyone, they argue, "can publish Music Ontology data and link it with existing data, in order to help create a music-related *web of data*." (Giasson and Raimond, 2007, Introduction)

They split their class and relationship definitions into three "levels" of complexity or "expressiveness": the first (least complex, least expressive) makes provision for describing common concepts in commercial recorded music such as "tracks / artists / releases". The second adds concepts for describing what they call the "workflow" of music production ("composition, arrangement, performance, recording"). The third provides a finer level of granularity in describing both music production processes (such as timelines) and musical details (such as pitch and harmony).

From FRBR it takes the four principal concepts *work*, *expression*, *manifestation*, and *item* and derives from them four equivalents for music: `MusicalWork`, `MusicalExpression`, `MusicalManifestation`, `MusicalItem`. It allows for the description of numerous classes of event in the lifecycle or production workflow of musical works including composition (often used as a noun correlating with a musical work, but here describing the act of composing a work), performance, recording, etc., and for relationships to be defined between those events and the objects (works, expressions, manifestations, items) and responsible parties (composers, performers, etc.) involved in the workflow.

In keeping with the Semantic Web principles of sharing concept definitions through linking, the Music Ontology makes use of several other network-delivered ontologies including some for describing time values and temporal events, for describing highly abstract properties of works, for describing persons, and for describing musical properties such as harmony.

The project to develop the Music Ontology is a work in progress at Queen Mary, University of London, directed by Frédérick Giasson and Yves Raimond. As a bibliographic encoding, it is potentially very rich in its detail. However, this potentially acts as an inhibitor to its uptake in that it requires considerable resources to generate such rich data. The success of the Semantic Web for music, at the same time, depends on a critical mass of data to make the linking of items meaningful.

## 4.8 Performative Encodings

The performative encodings are those which deal with musical data intended to be used either in real-time or at least in musical performance. As a result, like the framework encodings, they treat the musical data diachronically and have a significant and often central method of dealing with time and synchronisation.

Like the framework encodings, they often deal with musical data in an event-based manner, encoding things such as "note-lists". Often, the data is instructional rather than more typically musical. So instead of encoding pitch and time information, for example, they may encode instrument key and on-set and duration. Some are also more procedural than declarative, for example, they may consist of algorithms which generate sound.

They are often not very concerned with media or with sources and transcriptions as their intended dissemination is often in performance and is therefore transient and non-repeatable.

One advantage that this class of encodings often exhibit over others in this study is that, because of their application specificity, and because they are often procedural (i.e., part of an application program), they enjoy much more credible and creative use.

Some examples of performative encodings include:

- Waveform: various representations of the raw audio signal. Analogue (and non-computer) representations conform to the basic pattern of inscribing sound pressure on a medium moving at a constant rate. The digital corollary is to encode sound pressure *samples* as numerical values at such a rate (typically 44,100 per second) that the illusion of continuous sound is generated upon playback. Roads (1985, 405) describes waveform-like digital representations as "iconic" (as opposed to symbolic, employing semiotic terminology), arguing that "the patterns of number values mirror the patterns of the waveform."

- *MUSIC-N*: a series of music synthesis programs started by Max Matthews at Bell Labs with MUSIC I (1957). The project was developed in connection with research into telephony. From MUSIC III (1960) onwards the software was written in the programming language FORTRAN. The representation scheme consists of *orchestras* (collections of samples or instrument synthesis algorithms) and *scores* or *note-lists* which defined what pitches would be sounded and when using the available instruments. The MUSIC-N series is an ancestor of various current music synthesis (and live electronic performance) tools including Csound (which emphasises quality of synthesis algorithm), Common Lisp Music (which emphasises flexibility of pre-synthesis logic), and SuperCollider (which emphasises real-time use, experimentation, and performance).

- *UPIC* (*Unité Polyagogique Informatique du CEMAMu*): (Xenakis 1971–77) is essentially a composition system which consists of a graphics tablet onto which the composer draws waveforms and envelopes. The waveform images are stored and can be rendered sonically by the computer. The composer can then use the graphics tablet to arrange stored waveform objects in a 2D space (time against pitch) which includes abstractions such as "pages" (collections of waveform objects or events) and "scores" (collections of pages).

- *MIDI* (Musical Instrument Digital Interface): includes a representation scheme for describing instructions for digital instruments which is very much time-based (including a clock) and describes the actions to be performed on the instruments. (Described in detail below)

- *PML* (Performance Markup Language): a project being developed by Douglas McGilvray in an interdisciplinary group at the University of Glasgow. It is an XML-based format for encoding performance and analytical structures to be applied to existing XML-based music score formats such as MEI or MusicXML.

- *GDIF* (Gesture Description Interchange Format): (Jensenius, 2007, 205–226) is being developed by collaborating groups at the University of Oslo and McGill University as an attempt to define a format which handles movement data at various levels of abstraction and provides a secure synchronisation mechanism with other media data formats (including music notation). The format captures movement data from sensors and then encodes it for streaming in real-time. The encoded data may be streamed over the network and may use a multi-layered Open Sound Control protocol. As well as real-time streaming, the format also makes provision for applying analyses to the low-level data and encoding them as an additional layer in categories such as emotion, performance, environment, instrument and body. To aid processing of this data, the format will include an XML-based representation as well as the OSC-based one. The classes of quantitative data encoded in the multi-layered streams include velocity, segmentation, and trajectories, while the format also allows for more qualitative synchronised streams such as annotations.

### 4.8.1 MIDI

The Musical Instrument Digital Interface (MIDI) is several different things: a hardware interface to musical devices, a file format, the data in a MIDI file, or a standardised set of synthesised instrumental sounds. Its manifestation as a class of data specification is most relevant to this study, though it depends on an understanding of the hardware interface.

MIDI as a hardware interface consists of a protocol for sending musical events data between digital instruments and computers. Most fundamentally, that event data consists of just the key that was pressed on the instrument and the time it was pressed and the time it was released. Unlike many of the other encoding methods we examine here, MIDI (at least at this hardware interface level) bares no relation to the musical score, it is not a static document containing two-dimensionally arranged representations of musical information. Rather, it is continuous stream of data which (almost) only exists in the moment.

MIDI as a communication mechanism works in two directions: either as a recording process which listens for events from instruments and stores them in disk files, or as a broadcast process which reads events from disk files and sends them to instruments to render as sound. In any working MIDI system both these processes are controlled by a MIDI clock which timestamps events and regulates their dissemination. It is at this level that MIDI begins to qualify as a subject of this study. Both processes require the inscription of event data onto a digital medium.

Event data is stored in MIDI files in a binary (and therefore, importantly, non-human readable) format. Such files may organise their event data in one of three ways: 1) as a single *track* (`Format 0`); 2) as separate but temporally aligned tracks (`Format 1`); or 3) as separate and temporally independent tracks (`Format 2`). In each case the disposition of event data within the track is one-dimensional as its spatial arrangement only correlates with one musical dimension: time.

The data in the file includes a header which tells the processing equipment what kind of data to expect (number of tracks etc.) and the track data. The track data consists of a list of *time delta/event* pairs. Events may be one of three types: 1) MIDI events (note on, note off, etc.); 2) *meta-events* (track names, tempo indications, lyrics, etc.); and 3) *system-exclusive events* (events proprietary to particular equipment or software). Note events are described using five elements: 1) a time delta (or offset); 2) note on/note off status; 3) a *channel* (or track) number (each channel can be assigned a different instrument, so this property determines the timbral quality of the note event); 4) a note number (giving the pitch); and 5) an attack velocity (giving the dynamics). One important feature of this representational scheme to note is that MIDI events encode pitch as a numerical value (between 0–127 where 60 is middle C and each integer increment is a semi-tone). This is very different to diastemic notations such as CWMN where pitch is considered a dimension of music and so is represented spatially. Of course, as binary file formats don't make any attempt to circumvent the inherent sequential layout of disk files they are only capable of representing one dimension spatially, and in the case of MIDI that dimension is time.

It is partly for this reason that Wiggins et al. (1993) are critical of MIDI as a music representation method. Using their taxonomical classification system (described above) they describe it as neither very structurally general nor very expressively complete because it admits of no "multi-level" structural representations and employs a pitch representation scheme which is too abstract and can't, they argue, represent pitches outside of the equal temperament system.

As a binary format, MIDI is much less amenable to non-sequential (i.e. non-event) data and so it utilises meta-events not only to encode temporally arranged non-note data such as aligned lyrics and instrument changes, but also non-temporal data such as copyright notices and track names.

Time in MIDI is measured either as pulses per quarter note (PPQN) or using the Society for Motion Picture and Television Engineers (SMPTE) method which is based upon frame-rates for synchronising sound and video. In either case, the MIDI clock ticks continuously and triggers each event when the amount of time encoded it its time delta has passed. The temporal note events are simply `Note_On` and `Note_Off`, so no explicit concept of duration is represented in MIDI.

In some respects, MIDI actually displaces very little of the music. While notations (and their digital corollaries) employ spatial metaphors and force music to assume a two-dimensional nature which (particularly in the case of pitch) isn't an inherent feature, MIDI is arguably closer to the diachronic, linear nature of music in its representation of successive events with simple performative and sonic properties.

MIDI has, however, found itself a large audience. Support for it is ubiquitous amongst both sequencer applications and music notation editors. It generally affords a simplicity and reliability in application (particularly hardware applications) that make it very suitable for some uses in live performance.

# Conclusion

In this study we have seen how inscription provides a useful metaphor for how information may be stored in computers and what affordances computers may give for knowledge-based work. We have seen the importance of choosing suitable abstractions to deal with complex real-world concepts within the formal constraints of digital representations and of integrating a reasoning mechanism or semantics into any chosen representation. We have seen how collections of semantically rich inscriptions may be brought together into conceptual entities such as works and physical entities such as documents and those self-contained entities may be further collected and summarised in archives and catalogues. Throughout all, from the tiniest marks to the largest collections, the inscription metaphor as Latour describes it has been central.

However, as well as these purely representational issues, we have also attempted to say something about the communities and practice and sites of activity which make digital music encoding an interesting cultural practice.

The practice most directly related to the representational issues was the importance of modelling described by McCarty (2005), that the representation itself is less important than the process through which the representation was constructed, the intensive examination of the modelled phenomenon and repeated attempts at making a satisfactory model.

In arguing for the social construction of media histories, Gitelman (2006, 59) emphasises the importance of placing users and uses at the centre of media histories, rather than "product development, product placement, business models, or calculations of market share." She draws a distinction between "publics" and "users". Users are conscious users of a technology (or similar), publics are the intended or perceived (by the "inventors" or "producers") users, but users aren't necessarily in the public, and publics don't necessarily consist wholly of users, i.e. the producers may believe certain people to be in their target public who in fact aren't, and similarly, some people may consider themselves users who aren't in the producers' intended public. Users are "diverse, dynamic, and disaggregate". No one controls the definition of the community of practice.

The communities of practice which require digital music representations include those doing musicology with computers and those using computers in creative sound practices. The later group have largely provided for themselves and been quite successful in their design and application of suitable representational techniques (as we saw in our discussion of "performative" encoding practices). However, the musicologists still haven't taken up computational tools to any great extent.

Cook (2005, 7) argues that training in computational musicology should become commonplace. The skills he lists are: "basic principles of music representation, how to operationalize musicological problems, how to use different computational tools in conjunction with one another." He argues that, as an alternative, instead of training musicologists collaborations could be established between musicologists and computer scientists. He asks whether this would be possible if the musicologists didn't have sufficient understanding of operationalising musicological problems. "And how would such collaborations be funded? How would the music computing specialists develop their career routes? How would we avoid the very hierarchical relationships that currently characterize humanities research carried out by a 'principal investigator' in conjunction with 'research assistants'?".

These questions are bound up with the ideas beginning to emerge in new methodologies for musicology. Post-Kerman, musicology took on-board some of the postmodern scholarly practices which had been fashionable in other humanities disciplines since the 1960s, thus vastly increasing the scope of practices which came under the scrutiny of musicology. There is a feeling amongst current research students and even some professional musicologists, however, that these literature-based research methods have now been thoroughly investigated and should become more of a tempering device for new methodologies which draw on practice in other disciplines (particularly the sciences), than an end in their own right. As interdisciplinarity is currently very much in vogue, musicology is beginning to take advantage of the opportunities it has to offer by engaging practitioners from the sciences in studying musical phenomena. For example, cognitive scientists are contributing to our understanding of the nature of music, behavioural psychologists are helping us to formalise our descriptions of musical practice, and computer scientists are allowing us to deal with musicological problems on a scale larger than has even been possible before as well as providing tools and techniques to aide many other scholarly endeavours in musicology.

There are still hurdles to be overcome. How far are musicologists able and willing to accept the kinds of abstractions that computer scientists routinely deal with? Which aspects (if any) of contemporary musicological practice would suit algorithmic reduction? And what will be the cultural status of any results generated by such tools? Are they evidence of musicological practice or computer science practice? What division of responsibility is embodied in them? How should any ethical questions of attribution or acknowledgement of ICT-enabled techniques be addressed?

As an example, Cook (2005, 5) argues that many musicologists react against computational musicology on the grounds that musicological work is based on the experience of music and that computers remove that experience from the equation. He argues that this represents a "confusion between technology and epistemology". Cook also suggests that one possible answer to these problems is that computational musicology has to change to be seen as "doing musicology with computers", it has to become just another tool in the "musicological toolkit" and that it's status as something *other* or peripheral acts as a disbar to its uptake.

Finally, it's clear that these problems aren't new. Even as early as 1969, on the subject of building collections of encoded music, Barry S. Brook saw that "[t]he fundamental problem remains: collecting and counting, no matter how speedy our machines or how sophisticated our programs, will be no more meaningful than they ever have been without the application of intelligent hypotheses and accurate analytical techniques to valid, scholarly objectives." (Brook, 1969, 296)

# Bibliography


Abelson, H. and G. J. Sussman (1996). *Structure and Interpretation of Computer Programs* (2nd ed.). Cambridge, MA: MIT Press.

Babbitt, M. (1965). The use of computers in musicological research. *Perspectives of New Music 3*(2), 74–83.

Beynon-Davies, P. (1991). *Expert Database Systems: A Gentle Introduction.* London: McGraw-Hill.

Bowker, G. C. and S. L. Star (1999). *Sorting Things Out: Classification and its Consequences.* Cambridge, MA: MIT Press.

Brachman, R. J. and H. J. Levesque (2004). *Knowledge Representation and Reasoning.* New York: Elsevier.

Brinkman, A. (1990). *Pascal Programming for Music Research.* Chicago: University of Chicago Press.

Brook, B. S. (1969). Style and content analysis in music: The simplified "plaine and easie code". In G. Gerbner (Ed.), *The Analysis of Communication Content: Developments in Scientific Theories and Computer Techniques*, pp. 287?–296? New York: Wiley.

Brook, B. S. and M. Gould (1964). Notating music with ordinary typewriter characters (a plaine and easie code system for musicke). *Fontes Artis Musicae XI*, 142–155.

Brown, C. H., J. Kolar, B. J. Torrey, T. Truöong-Quang, and P. Volkman (1976, February). Some general principles of biological and non-biological folk classification. *American Ethnologist 3*(1), 73–85.

Butterfield, M. (2002). The musical object revisited. *Music Analysis 22*(3), 327–380.

Byrd, D. and T. Crawford (2002). Problems of music information retrieval in the real world. *Information Processing and Management 38*, 249–272.

Camilleri, L. (1992). Computational theories of music: Theoretical and applicative issues. In A. Marsden and A. Pople (Eds.), *Computer Representations and Models in Music*, pp. 171–185. London: Academic Press.

Chandler, D. (2007/2002). *Semiotics: The Basics* (2nd ed.). Routledge.

Cole, H. (1974). *Sounds and Signs: Aspects of Musical Notation.* London: Oxford University Press.

Cook, N. (2001). Between process and product: Music and/as performance. *Music Theory Online 7*(2), 1–31.

Cook, N. (2005, September). Towards the compleat musicologist? In *6th International Conference on Music Information Retrieval*, London, UK.

Coward, R. and J. Ellis (1977). *Language and Materialism: Developments in Semiology and the Theory of the Subject.* London: Routlegde.

Crawford, T. (1991). Tabcode for lute repositories. In E. Selfridge-Field and W. Hewlett (Eds.), *Computing in Musicology 7*, pp. 57–59. Cambridge, MA: MIT Press.

Dahlström, M. (2006). *Under utgivning: Den vetenskapliga utgivningens bibliografiska funktion.* Ph. D. thesis, University of Bor, Bor, Sweden.

Dale, N. and J. Lewis (2006). *Computer Science Illuminated* (3rd ed.). Jones Bartlett.

DeRose, S. (2004, August). Markup overlap: A review and a horse. In *Extreme Markup Languages*, Montréal, Québec. http://xml.coverpages.org/DeRoseEML2004.pdf Accessed 10 May 2009.

Dodge, B. (2002). Across the great divide: Archival discourse and the (re)presentations of the past in late-modern society. *Archivaria 53*(Spring), 16–30.

Erickson, R. (1976). Darms—a reference manual. Technical report, Queen's College, New York.

Floridi, L. (1999). *Philosophy and Computing: An Introduction.* London: Routledge.

Geertz, C. (1993/1973). *The Interpretation of Cultures: Selected Essays.* London: Fontana Press.

Giasson, F. and Y. Raimond (2007). Music ontology specification. http://musicontology.com Accessed 10 May 2009.

Gieseking, M. and T. Weyde (2002). Concepts of the musitech infrastructure for internet-based interactive musical applications. In *Proceedings of the First International Symposium on Cyber Worlds (CW'02)*, Volume 00, Los Alamitos, CA, USA, pp. 30. IEEE Computer Society.

Gitelman, L. (2006). *Always Already New: Media, History, and the Data of Culture.* Cambridge, MA: MIT Press.

Goehr, L. (1992). *The Imaginary Museum of Musical Works.* Oxford: Oxford University Press.

Goldfarb, C. F. (1984). The standard generalized markup language: Basic concepts. In F. Krückeberg, S. Schindler, and O. Spaniol (Eds.), *Offene Multifunktionale Büroarbeitsplätze und Bildschirmtext*, Informatik-Fachberichte, pp. 132–140. Springer.

Goodman, N. (1968/1976). *Languages of Art* (2nd ed.). Cambridge, MA: Hackett.

Grammit, D. (1998). Musicology, commodity structure, and musical practice. In P. F. Broman, N. A. Engebretsen, and B. Alphonce (Eds.), *Crosscurrents and Counterpoints: Offerings in Honor or Bengt Hambraeus at 70.* Göteborg: University of Gothenburg, Department of Musicology.

Hall, T. (1997). Darms: The a-r dialect. In E. Selfridge-Field (Ed.), *Beyond MIDI: The Handbook of Musical Codes*, pp. 193–200. Cambridge, MA: MIT Press.

Harold, E. R. and W. S. Means (2004). *XML in a Nutshell* (3rd ed.). O'Reilly.

Howard, J. (1997). Plaine and easie code: A code for music bibliography. In E. Selfridge-Field (Ed.), *Beyond MIDI: The Handbook of Musical Codes*, pp. 362–372. Cambridge, MA: MIT Press.

Huron, D. (1992). Design principles in computer-based music representation. In A. Marsden and A. Pople (Eds.), *Computer Representations and Models in Music*, pp. 5–39. London: Academic Press.

IFLA (1998). Functional requirements for bibliographic records. Technical report, IFLA, The Hague. http://archive.ifla.org/VII/s13/frbr/frbr1.htm Accessed 10 May 2009.

Impett, J. (2000). *Computational Models for Interactive Composition / Performance Systems*. Ph. D. thesis, Cambridge University, Cambridge.

Jensenius, A. R. (2007). *Action - Sound: Developing Methods and Tools to Study Music-related Body Movement*. Ph. D. thesis, Department of Musicology, University of Oslo, Norway.

Kerman, J. (1985). *Musicology*. London: Fontana Press.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.

Lakoff, G. and M. Johnson (1980). *Metaphors We Live By*. Chicago: University of Chicago Press.

Latour, B. (1985/1990). Visualisation and cognition: Drawing things together. In M. Lynch and S. Woolgar (Eds.), *Representation in Scientific Practice*, pp. 19–68. Cambridge, MA: MIT Press. http://www.bruno-latour.fr/articles/article/21-DRAWING-THINGS-TOGETHER.pdf Accessed 10 May 2009. Page numbers cited refer to online edition.

Lerdahl, F. and R. Jackendoff (1983). *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.

Levy, D. M. (2001). *Scrolling Forward: Making Sense of Documents in the Digital Age*. New York: Arcade Publishing.

Lomax, A. (1976). *Cantometrics: A Method in Musical Anthropology*. Berkeley: Extension Media Center, University of California.

Lunenfeld, P. (1999). Unfinished business. In P. Lunenfeld (Ed.), *The Digital Dialect: New Essays on New Media*, pp. 7–22. Cambridge, MA: MIT Press.

McCarty, W. (2005). *Humanities Computing*. Basingstoke: Palgrave.

McClary, S. (1993). Narrative agendas in "absolute" music: Identity and difference in brahm's third symphony. In R. Solie (Ed.), *Musicology and Difference: Gender and Sexuality in Music Scholarship*. Berkeley: University of California Press.

McNeill, W. (1992). *The Pursuit of Power, Technology, Armed Forces and Society Since A. D. 1000*. Chicago: University of Chicago Press.

Negroponte, N. (1995). *Being Digital*. New York: Alfred A. Knopf.

Newell, A. and H. Simon (1976, March). Computer science as empirical inquiry. *Communications of the ACM 19*, 113–126.

Nunberg, G. (1979). The non-uniqueness of semantic solutions: Polysemy. *Linguistics and Philosophy 3*, 143–184.

Reynolds, R. (2005). *Mind Models: New Forms of Musical Experience* (2nd ed.). London: Routledge.

Roads, C. (1985). Grammars as representations for music. In C. Roads and J. Strawn (Eds.), *Foundations of Computer Music*, pp. 403–442. Cambridge, MA: MIT Press.

Roland, P. (2002, September). The music encoding initiative. In *Proceedings of the First International Conference on Musical Applications using XML*, Milan, Italy.

Roland, P. (2006). The music encoding initiative (mei) dtd and the ocve. Technical report, OCVE. http://www.lib.virginia.edu/digital/resndev/mei/mei_ocve.pdf Accessed 10 May 2009.

Ruwet, N. (1977). Theorie et methodes dans les etudes musicales: Quelques remarques retrospectives et preliminaires. *Musique en Jeu 17*.

Searle, J. (1980). Intrinsic intentionality. *Behavioral and Brain Sciences 3*, 450–456.

Selfridge-Field, E. (1997). Describing musical information. In E. Selfridge-Field (Ed.), *Beyond MIDI: The Handbook of Musical Codes*, pp. 3–38. Cambridge, MA: MIT Press.

Simon, H. (1981/1969). *The Sciences of the Artificial* (2nd ed.). Cambridge, MA: MIT Press.

Skinner, R. (1990, March). Music software. *Notes 46*(3), 660–684.

Small, C. (1998). *Musicking: The Meanings of Performing and Listening*. Hanover, NH: Wesleyan University Press.

Smiraglia, R. P. (2001). Musical works as information retrieval entities: Epistemological perspectives. In *Proceedings of the 2nd Annual International Symposium on Music Information Retrieval*.

Thompson, H. and D. McKelvie (1997). Hyperlink semantics for standoff markup of read-only documents. In *Proceedings of SGML Europe'97*. http://www.ltg.ed.ac.uk/~ht/sgmleu97.html Accessed 10 May 2009.

Weinberger, D. (2002). *Small Pieces Loosely Joined: How the Web Shows us Who We Really Are*. Oxford: Perseus.

West, R., P. Howell, and I. Cross (1991). Musical structure and knowledge representation. In *Representing Musical Structure*, pp. 1–30. London: Academic Press.

Wiggins, G. A. (2009). Computer-representation of music in the research environment. In T. T. Crawford and L. Gibson (Eds.), *Modern Methods for Musicology: Prospects, Proposals and Realities*, Digital Research in the Arts and Humanities. Aldershot, UK: Ashgate. Preprint available at http://www.methodsnetwork.ac.uk/redist/pdf/wiggins.pdf Accessed 10 May 2009.

Wiggins, G. A., A. Smaill, M. Harris, and E. R. Mirana (1993). Surveying musical representation systems: A framework for evaluation. *Computer Music Journal 17*(3), 31–42.

Winston, P. H. (1984). *Artificial Intelligence*. Addison-Wesley.