

s o n i c c o n t r o l
i n
INTERACTIVE AUDIO INSTALLATIONS

thesis submitted
8 SEPTEMBER 2005
for

SECOND PHASE SONOLOGY
at
KONINKLIJK CONSERVATORIUM
DEN HAAG

by
AMY BEESTON
amy@cunningbees.co.uk

chapter one

Experiences in the fields of music and technology have led me to become interested in the manner by which humans and machines can, together, deal with a sound picture. New synthesis techniques to produce electronic music have become well developed in recent years, but methods of sonic control have tended to be more fully implemented in areas outside of music research, for example in tasks such as voice transcription. I am primarily researching machine listening models in order that I may better enable the computer to filter and absorb 'relevant' information from the surrounding audio environment. Alongside the development of these machine-listening tools, many sound-producing patches have also emerged during my studies in Sonology, ready and keen to exploit the side-products of these analysis units. Signals exist which, although designed for 'control' purposes, may also be listened to 'raw' or as subsequently processed audio signals. I am currently experimenting with these signals as sound material, in composition, performance and installation settings, and will describe some results in this document.

The main context for this thesis will be audio installation art, maintaining an open mind and taking a broad survey of those things often referred to as responsive environments, audio environments, interactive installations, site-specific works, etc. These works form a movement that seeks to address the questions surrounding the crumbling distinction between performance and composition. I will discuss works designed by others and by myself that explore varied contexts of aural-presentation from more formal, traditional settings such as the concert hall and gallery space, to some alternative venues which now host these artworks such as the club and cafe. The installations to which I will refer sometimes rely on listening models which enable the computer to gain sound information from the situation experienced in the immediate environment.

Along with sound, other media are inevitably involved in every live presentational situation, whether it be the case of a human being performing on an acoustical instrument, sound delivered by exclusively electronic means, or a combination of both. There are fundamental choices to be considered with respect to the audio environment - room acoustics, equipment, etc. - right through to other realms of sensed variables, for

example offering variable lighting conditions in the space.

Involving a conglomeration of many different disciplines, this work is often collaborative. It is interesting to me in that respect as it prompts discussion of all aspects of imagination and intention, bluntly exposing the necessity to develop our methods of communication alongside our art - the actual works of non-verbal format, be it a composition for a fixed-media tape piece, practice for an instrumental performance or indeed audio interaction modeling for an installation. Concepts commonly shared will be outlined and discussed, trying to ascertain their aesthetic, cultural and social implications. The idea of communication - passing messages - is central to a discussion of interaction, whether or not the situation is of the audio art-work variety. People and machines can be combined in endlessly re-figured modes of collaboration and individual cases will be discussed here.

With an open mind, the term 'interactive' can help to raise questions and open possibilities for understanding the things we do, the tools we use, the sounds we make and the music we compose and play. We must remember that **"through history, all art asked for audience interaction"** [Wilson 2002 p655], however the very word 'interaction' often conjures negative images in the minds of electronic composers today. This fact is largely due to the usage of the word in the commercial world, where particularly in relation to video games, we find - **"the term "interactive music" describes systems that make music without any musical input from the user at all, providing audio that changes in synchrony with game-playing - a byproduct of joystick movement, for example"** [Rowe 2001 p379]. This thesis will address this issue in some detail, promoting the opinion that interactivity and audience involvement is a positive goal to aim for: it requires one to **"think and act like a creative person with intelligence, imagination, and musical expressivity"** [Chadabe 2002 p2].

I am considering the processes by which digital media can, through the medium of sound, understand a human being or a traditional (but none-the-less technological) instrument. Rather than conceive of the instrument purely in the sense of a tool that can be manipulated to achieve certain tasks, we find that with a computer-based electronic set-up it is often the case that **"the link between a controller and sound generator can be active, which is to say that the**

link can be a computer that can generate information, run algorithms, or act in any way as an intermediary between a performer's controls and the variables of a sound generator" [Chadabe 2002 p1]. This thesis examines work undertaken into movement detection and sensing, parameter-mapping strategies, and the concerns of gestural expressivity regarding the human-computer interface. It then turns in detail to the most immediate musical aspect of the research-performance-composition process, that is, the sound.

The very nature of this medium is fluidity, and as such, it would be difficult, even unhelpful, to chart a linear flow through the various thinking processes behind such works. The reader will therefore find important ideas recurring repeatedly, recombined with and encircling one another, re-interpreted from a different perspective in each encounter in order that the surface beneath the initial scratch be examined far more closely.

chapter two

In his history of sound in the arts, Douglas Kahn states that Cage used chance and indeterminacy as compositional techniques **"not only to eliminate himself from his music but to eliminate the social situations in which he found himself"** [Kahn 1999 p189]. The infamous 4'33" was held in a concert setting **"where any muttering or clearing one's throat, let alone heckling, was a breach of decorum. Thus, there was already in place ... a culturally specific mandate to be silent, a mandate regulating the behavior that precedes, accompanies, and exceeds musical performance"** [Kahn 1999 pp165-166].

A large amount of confusion exists surrounding interactive installation art. What is an installation? What is interactive? Talking about interactivity in art leads us straight into troublesome territory but provides an excellent playing field for **"exploring what it means to be embodied in high-tech worlds"** [Harraway 2000 p310]. Later in this thesis I will consider methods of listening used to make sonic installation artworks. This constitutes a relatively small body of work under discussion so we can afford to examine the enveloping situation in some detail.

Many audio installations respond to a human being by playing music, but this so-called interaction is commonly achieved by non-sonic means. The concept can be introduced with a simple idea seen on a garden show on television: any time you opened the back door and walked outside, a motion sensor was triggered and music would begin to play. Here, the state of the door is measured: open or shut, as a binary logical decision. When the door opens, a CD plays, beginning a new track at random if no sound is currently being produced.

We live in a visually dominated society, receiving proportionally far more information through our eyes than our ears. Participation and movement, however, are both inhibited by the 'sit-down, eyes-front' listening culture that has become established in the traditional concert setting of the western art world. Interactive art, on the other hand, makes it obvious to us that we are a part of some integrated system, no longer the disinterested (perhaps even passive) auditor/viewer. Rather, an interactive situation becomes a self-reflexive work in that it is, on some level, created by and for each individual interactor. Interactivity **"exists between a creator and a person who experiences a finished work"** [Ritter 1996]. In this vein,

"Rokeby sees interactive installations as micocosms in which viewers can assume responsibility for their actions and reflect on their status in larger social systems" [Wilson 2002 p654].

The installation artist Don Ritter points out that, in discussions of art, a medium usually refers to **"a physical substance that carries impressions to the human senses"** [Ritter 1996]. However, in installation art we now use the term interactivity to describe the various methods available to interact with different media, referring to the method of communication between people and media, judging how one can put in information compared to how one perceives the output. Interactivity is not a physical substance, is not a medium in itself.

Installations involve different domains of work. Installation artists have many of the same things to deal with that are met in traditional studio methods of tape composition, but must also consider extra 'data' relating to the presentational situation. Interaction with sound should be promoted wherever the work is to be dealt with in a primarily aural perceptual ground. This occurs rarely in audio installation art, however. Rather, audio output is commonly governed by one or more of audio input, video input, controller input, sensor input etc., which in turn may be inputted by man or machine. Installation artists must therefore make the languages meet, make the devices relate, make the music happen.

Installations frequently involve projection of video or animation sequences. A film is primarily visual, with the two-dimensional flat screen presenting the show. The sound-track, including the audio backdrop and spoken dialogue, provides another dimension to support the visual action and usually (especially in narrative film) aims to reinforce the feeling of a particular reality in the viewer, that they are really there. This trick for increasing one's immersion is not required, however, in an installation environment: the person really is there, so the sound is freed from its supporting role and is allowed to stand alongside any visual element as primary action. Rather than considering the visitor as a 'viewer' with a fixed line of sight forwards, the 'listener' in an installation is able to rotate their head, to move around the room. Further, in an interactive installation, there will frequently be a reward for participation, as is found in the interaction within the games world where specific tasks are

designed to promote skill-directed learning techniques.

It is often useful to coordinate sound and visual programming in one and the same software environment, giving a visual display of the events that take place in the audio within the room as they are understood by the system computer. A number of analysis techniques have therefore been developed to enable a visual-display feedback of the real situation as it is represented in the machine environment. Some of these methods will be described later in this thesis. These methods allow the artist to easily compose changes in displayed imagery alongside some variation in input signal and constitute an element found very commonly in interactive installations, giving a similar reward to those in games.

Although this thesis does not focus explicitly on the combination of audio and video domains of (re)presentation, it is worth reminding the reader that **"the true potential of a radically asynchronous sound film has, to this day, not been adequately explored"** [Kahn 1999 p124]. In film, the interaction of sound with image can be extensively varied (a/synchronous presentation, sound and visual time reversal, etc.) in order to stimulate different perceptions in the viewer. Many artists are currently working in this paradigm. With audio, when a sound is presented through loudspeakers the source is separated from its direct acoustic source. Recording has thus brought about the ability to reproduce copies without the need for an original to be present, and these can of course be re-processed and re-presented in many different ways. The situation becomes more complicated still in the case of composing for interactive multimedia environments, where one faces the problem that **"the cross-modal interpretation of musical information can be difficult to make clear to an audience"** [Rowe 2001 p317].

Another potentiality that confronts the installation artist is the gradual breakdown of the boundary between humans and machines in recent history. Donna Haraway refers to the **"leaky distinction"** between animal-human (organisms) and machine, pointing out that there is a **"very imprecise"** border between the physical and non-physical [Haraway 2000 p293].

In this fashion, Toshio Iwai is an inventor of installations that **"blur the boundaries between the digital and the physical in several ways. Actions taken in the digital world**

activate physical objects, and physical objects control digital informations" [Wilson 2002 p766]. In such installations it becomes increasingly important to question whether the human is telling the machine how to act or whether the machine is in fact prompting the human as to how to act. Is the human the master of his tools or is it the instrument itself that is controlling the situation? Is there something essentially musical about being able to blow, strike or rub an instrument, other than actually being able to control one's repeated movements and to be able to do a specific action in a specific timeframe and according to certain culturally established traditions? Perhaps this is what our machines do, too?

'Play' in an instrumental sense leads to a discussion of the expressiveness of interfaces regarding the content of messages passed onwards, and also to the possibilities of extracting emotional data from expressive gestures. In the realm of installation art the concept of 'play' differs: it addresses the cross-modal nature of the installation and should question how to **"make the relationship between visual, audible, and physical gestures meaningful"** [Rowe 2001 p355]. Robert Rowe continues, **"when watching a performance, observers deduce relationships between the humans onstage and their machine partners. When faced with an installation, observers must explore the nature of the relationships by interacting with the environment themselves"** [Rowe 2001 p355].

Stephen Wilson makes plain many of the common strategies of interactivity throughout his book 'Information Arts' [Wilson 2002]. Rather than open up real worlds of new activity to the participant, the installation artist frequently sets up a charade of choice regarding the timing and type of access to options bringing about change in the environment. Presence, simple choice, choice from a set of options, or rigid sequential structures are frequently offered rather than a structure with flexible timing of open choices which would allow more freedom of expression to the visitor. He states **"the mere act of making choices does not necessarily result in significant artistic interchange. Also, the interactive paraphernalia of computing ... cannot be separated from the history and conventions and social niches of computer use in the mainstream. Conventional interactivity comes out of the disciplines of computer-human interface design and engineering, whose agendas focus on efficiency and productivity rather than on more artistic goals such as provocations, discovery, nuance, and exploration"** [Wilson 2002 pp653-654]. The situation then becomes one of balancing the visitor's acts of creation and consumerism, their role as contributor or author and the nature of the interaction that can bring

these situations about.

This text is primarily concerned with interactions using audio signals, focused on the possibilities of using information derived from an audio signal for interaction within an installation environment. This topic returns more explicitly in later chapters. For now it suffices to draw the reader's attention to the fact that enabling audio trigger control points, by targeting specific audio events, is quite different to the use of the audio signal for processing audible manipulations in sound only. Questions arise as to what elements in an audio signal could be the reason for an interaction to occur, and what type of interaction this should be.

In music, even in life, we are certainly sensitive to sudden changes in volume, for instance, and therefore 'becoming suddenly loud or suddenly quiet' is a feature that we can look out for in the sound. Aside from our current standards of measuring 44100 amplitude points per second in air pressure fluctuation, there are many ways to describe what happens in sound, but, like movement patterns of instruments or other bodies, the vibrational patterns in air likewise follow the laws of physical matter. These laws dictate, for example, that dense substances conduct sound better than sparse ones. Something whose local time-scale average level increases could be said to be becoming more highly wound, more tense whereas something whose average local time-scale level decreases could be said to be unwinding, relaxing. A high value measured compared to a low one suggests something is more, say, brittle than floppy.

Here, we immediately glimpse the problem faced by experimental researchers in the 1970s who, in an attempt to clarify human perception of sound timbre, were limited in their work by linguistic association and semantic meaning of the actual words used in the analysis method. Although some researchers now do work with sound as their primary concern, it is more common for analytical work to be done on physical action to produce sound. In their research on gestural control of sound synthesis and processing algorithms, Arfib and Kessous [2003] worked out an analysis of movement data in reverse from the musical intention and sonic production method behind the sound. At this stage we can note that it is helpful, when trying to describe our experiences, to seek methods that can spread perceptual

clues across the various dimensions of analysis under consideration. Later chapters show the processes of 'looking' at various signals, describing the implications of programmed implementations and implied consequences of matching techniques etc. that are used in models of listening or playing.

In terms of precise variables, one only finds what one is looking for. In order to look for many things, it is therefore instructive to have multi-dimensional analysis methods that can be put in place in multi-modal environments that **"adapt their behavior to interaction with a human partner or other machine agents. The system extracts information about the actions of other agents in the environment (human or machine) by reading a collection of motion [or other] sensors"** [Rowe 2001 p373]. As one must learn to communicate with machines in their languages, one must learn to communicate with other artists and researchers in order to share thoughts, integrate the various modes of investigation and thus consolidate musical forms and structures.

Di Scipio's understanding of interaction as **"a network of interdependencies among system components, and as a means for dynamical behaviour to emerge upon the context of an autonomous system ... with the external environment (room or else hosting the performance)"** [Di Scipio p269] suggests an attempt to link otherwise separate dimensions. This is in an attempt to find interesting environments to resonate the computational model in a particular setting. In the same way that a location in a city can be coordinated by a grid-reference number, we can imagine that places in an N-Dimensional space can be given reference points too, most likely in a coordinate system of values in arrays of labeling points. Time-dependent vectors can be included that show change, specifically in how that place was arrived at, and the potential futures can be similarly transmitted. By analogy, the system is a long corridor (potentially infinite) along which one can find doors and access specific areas of sound and control settings. Composition thus describes the potential to make journeys through sound, with the data from input representations being continually mapped against templates which may or may not lead to side rooms off this corridor.

In order to construct a particular experience through potential events, aiming to re-combine dimensions and to articulate transitions between them, we can look to a method such as Insook Choi's manifold interface. This is

"an exploratory tool rather than a display tool" [Choi 1997 p17] that enables an active public to participate consciously in their own musical experience. By analogy with eating habits of today (where we find that many people simply do not cook and instead survive almost wholly from heavily processed ready-meals) we also find that many people simply consume whatever music is easy for them to reach. Rather than follow this trend, I would promote the alternative view, that, **"if interactive music systems are sufficiently engaging as partners, they may encourage people to make music at whatever level they can. ... it is critical to the vitality and viability of music in our culture that significant numbers of people continue (or begin) to engage in active music making, rather than simply absorbing reproduced music bombarding them from loudspeakers on every side"** [Rowe 2001 p4].

In a sense, this is an art-form that trades on the confusion of time in the present age. We see a trade-off: we have gained an immensely liberating set of working materials and have the potential to do many great things with them, but there is as yet little understanding of what constitutes a great thing, or of how to bring it about. Electronic arts and digital music production point towards the future and, according to the long tradition of music development in all cultures through the ages, are oriented towards the anticipation of new 'historical' situations. At this juncture, when many technologies are developed for practices of warfare and destruction, the installation art world provides a realm in which to be more positive, to build hope, to prepare for times when humans can 'naturally' have a far more intimate relationship (communication method) with machines than we do at present.

chapter three

This chapter of the thesis addresses the distinction between composition and performance, situating these concepts within the field of installation art with interactive electronic media. These installations deal with live sound, but also with non-live sound. The same is frequently true for live performance of electronic music. In either the case of performance or installation art, it is likely that some of the musical materials heard in the occasion have been preconceived and that some are artifacts of the actual presentational setting of the sound, perhaps in the musician's interpretation of instructions during the performance or in the actions of a participant in a so-called responsive environment.

The idea of a performer as an integral part of performance is ingrained. Without there being such a person present to perform, one can at once feel that there is something missing in a performance situation. Placing loudspeakers alone on a stage can give rise to this feeling, especially in the uninitiated listener. Electronic instruments differ from acoustic instruments by one main feature, namely that the the acts of sound control and sound generation are separate. That is, the sound itself has become distinct from the ways by which it can be altered, that the sound source and control interface are separate. Much research has been done on synthesis techniques since simple sine-wave generators moved within the composer's grasp, but far less has been learned about the methods by which these sounds may be controlled, far less on how they may be 'played' with a certain musical intention.

Electro-acoustic composers, seduced by new possible extensions of sound worlds undiscovered, unearthed, made up from scratch or offered by divine inspiration, have in general busied themselves with work in structures of sound that do not rely on the motoric and cognitive performance precision abilities of a human body. Rather, in a manner akin to the old masters, the composers of previous centuries, they coordinate the sound output of various instruments through time (often played by plain old electricity rather than by people). Whatever the music, techniques are passed on in some manner (whether by written scores and particular teaching methods as in the western art music background, or by one of the many unwritten traditions passed on by aural imitation in other parts of the globe). Each

method maintains some similarities and creates new variations. Composers of this century are still performing the same 'sound-coordination' tasks, but now there are still more options laid out before us - the new media, available to us can be moulding into many different forms and introduce ideas to music that have not been met before in this context.

Whether considering an acoustic instrumental performance on a stage in a traditional concert-hall setting of performance that has built up through countless generations of Western thinking in the art world, or whether considering a participant's actions in an experimental installation environment, we now find that technologies **"have recently textualized our bodies as code problems on the grid of C3I"** (command-control-communication-intelligence) [Harraway 2000 p311].

Joel Chadabe states plainly that **"as a general rule, the most important requirement of an instrument for a professional performer is that the instrument demonstrates for the audience that the performer is necessary and is in fact controlling the music ... Such requirements of obviousness do not exist when the instrument's behavior need be understood only by its performer ... the creativity in the instrument's output can be in itself rewarding"** [Chadabe 2002 p3]. This highlights the point that there is still reason to discuss control interfaces from a physical level right through to the possibilities of 'tracking' the player by invisible detection methods in audio or video.

It remains to be seen **"whether the 'instrument' metaphor is entirely useful when discussing interactive music systems"** [Di Scipio 2003 p269], but greater discussion is certainly to be encouraged. Further chapters discuss the concept of mapping and **"challenge the assumption that an electronic instrument consists solely of an interface and a sound generator"** [Hunt et al. 2002 p1] but also point out that **"mapping is a less useful concept when applied to the structure of complex and interactive instruments in which algorithms generate control information"** [Chadabe 2002 p1]. However, it is useful first to look at this situation from a more traditionally musicological standpoint, and to see some of the effects that the development in technologies is having on the music being produced.

Addressing the question of elements used as compositional techniques to structure longer sections of time, we see that Western classical music has been concerned primarily with balance [Straus 2004 p20]. Early formal

structures in music were related to the music's function (eg. the sarabande which originated in Spain as a courtly mating ritual) and evolved through actually playing the music rather than acting analytically according to the bodily and temporal requirements of the situation. 'Form' could be said to be an important element of musical thinking around the mid nineteenth century, however, and the classical tradition reflects the balance-imbalance-balance template in the sonata form's unfolding exposition-development-recapitulation. Prior to this, harmony was used to point out structural changes in form (eg. cadencing at the edges of phrases), and before that, melody. Through the gradual establishment of the 12-tone equal temperament convention, tonality was then used to bring about changes in the harmonic function of chords. Counterpoint, an idea of fusing at least two things together into a larger framework that allows both to exist as independent ideas and yet to stand in relation to one another, is perhaps fundamentally underlying all these developments as it introduced the complexity that allowed us to consider time and frequency as separate but simultaneous 'horizontal and vertical' dimensions rather than relying on a single dimension of synchronous harmonic change.

Outside of this period of development in the western art world, tonal and harmonic concerns are of lesser importance in the structuring of music. Most pop, rock, dub and reggae music will use just a few chords, frequently moulded in basic poetic forms such as alteration of verse and refrain. Rhythmic elements may come more to the foreground of attention in some styles (drum and bass etc.), but more often there is a vocal line to transmit a message into the listener's consciousness.

When we grant the fact that music may be produced live by humans on stage, or presented live over loud-speakers on stage, we enter different contexts. When a human being is performing they require a certain set of conditions in order to concentrate on the task in hand, allowing them to cope with the task of presenting the music they are prepared to play. Traditionally, a musician on stage expects silence from the audience while a background musician will expect people to be chattering through their performance. As a human, it can be difficult to deal with an unexpected response or unexpected event such as a mobile phone starting to ring. However, we cannot entertain the notion that a loudspeaker has expectations, only that

the person who is controlling it does. Therefore, when music is presented solely by a machine there can be no situation that would cause a 'lapse of concentration'. When there is an interruption in the "normal behaviour" of the music-playing-machine, it is usually interpreted as a technical hitch or malfunction of some description.

Since a machine of one kind or another that provides music has now entered almost every home with the advent of electricity, we have the ability now to listen to increasingly greater repertoires of music than our parents' generation. They in turn had wider access to music than their parents. Player-pianos could automatically reconstruct performances of popular works, and organ recitals in local town halls would allow reductions of symphonies to be heard. People now have control of what to listen to without having to learn an instrument themselves, without composing, without gathering other musicians together, and without going through any process of rehearsal.

Some concepts used as compositional constraints by composers today, however, are still familiar to the non-trained musician through the very pieces of technology that deliver sound to them. For example, bass and treble frequency equalization 'tone' controls are feature on some amplifiers, and alternative environments such as for a stadium or cathedral can be applied on others. Indeterminacy and chance, which entered music in the form of aleatoric and algorithmic composition, can be seen on many CD players where there is usually an option to randomize the tracks or re-program them in a specific order. Indeed, software packages such as iTunes allow many different re-orderings of the whole library of music or subset playlists thereof. Some basic equalization is also available, with display of the spectral content of the sound and an option to associate a particular acoustic processing with a particular track. Vinyl record players offered a choice of play rate. Whilst it was originally intended that this be set once per disk, this feature is now exploited by the DJ who can re-spin or scratch varied re-orderings across a side of vinyl or indeed across multiple decks.

Perhaps music involving electricity has to be concerned, at some level, with concepts of storage, retrieval and reproduction. The influence of

recent technologies can be heard very clearly in much of the 'intelligent dance music' produced in recent years, making use of both the functions and malfunctions of various techniques of sound manipulation. Short repeating loops, playing on the manner in which CDs 'skip' when damaged, can be found for example in the music of Fischerspooner and Fennesz. The notational score is useful at present neither for describing nor recreating much of this music, and a text print-out of the software programming environment will often be more instructive. Before returning to the outlined concepts of performance and composition, I would like at this stage to return to the interactive audio installation environment and discuss two installations in particular.

The 'Memory Machine' is an installed work presented in London in 2002, first at Cybersonica and later at the British Museum (in 2003). After answering a few questions by pressing a big button between the speakers, the creators (Nye Perry and Cathy Lane) allowed the interactors to record spoken tales of their experiences on specific topics (dreams, people, places etc). The resulting audio recordings would then be manipulated according to the topic category selected, and according to the interactor's decade of birth. The environment would mix this new memory in with others already held in the machine, and provided a sound-scape with vocal fragments of linguistic sense alongside musical elements directed by the categories of selection that had been previously associated with the recording.

A very similar work was created in collaboration between Iain Mott and Marc Raszewski and premiered at the Art In Output Festival, Eindhoven 2003. "Summoned Voices" uses the metaphor of a door-buzzer intercom system, prompting interactors to press a button and introduce themselves by name or otherwise in words, song, noise etc. Again, this sound is mixed with that of previous users and re-composed into an audible vocal sound-scape. The creators' belief is that **"interactivity heightens the experience, engaging and directing participants in an activity of sensory exploration. Participants play a focal role, lending a great deal of creative input. Ultimately outcomes are a result of the partnership between the artists and the public"** [Mott and Razewski 2005].

These installations provide a good example of how electronic music enjoys a

freedom not so easily afforded to other musicians - primarily that time and re-production techniques can be controlled in computational systems, making use of delays, buffers, and many other digital signal processes to reset, redo, reassess, and remake the audio signals from one sound object. Many sound artists are reworking the same ideas, using the same processes, the same digital implementations, with particular musical intentions. Ideas often loop and re-loop amongst friends and colleagues sharing conversation and ways to make music in a collaborative process which will be discussed more thoroughly in Chapter four.

The separation of sound source and its mechanisms of control has left us with the impression that things can be set up in an infinity of different ways. Our challenge has been to match these possible ways of organization with those necessary for the particular context of presentation. This simple factor (a difference in physical space) prompts us to question what aspects may be deemed compositional or performative about each work. Robert Rowe highlights two alternative approaches to this situation where music is produced by humans and machines in combination.

"Instrument paradigm systems are those that treat the machine contribution as an extension or augmentation of the human performance. Player paradigm systems present the machine as an interlocutor - another musical presence in the texture that has weight and independence distinguishing it from its human counterpart" [Rowe 2001 p302].

When we look back in history at the devices that have been used to produce sound, we find that after much experimentation **"those that became accepted into the repertoire came to be called the standard instruments"** [Wilson 2002 p407]. When playing an acoustic instrument the musician manipulates the mechanical action of the instrument, and in so doing controls the parameters of the instrumental timbre directly. A violinist bows the very same strings that transfer vibrations through the bridge to the back plate of the body before radiating sound energy to the world. A trumpeter blows (via a mouthpiece, itself a cupped section of a resonating metal tube) into the same tubes that eventually transmit sound from the bell of the instrument. The trumpet maker therefore must consider the physical laws that govern this mechanism, for example that higher frequency waves spread out less quickly than lower frequency ones at boundary junctions. The conical shape of the

flare thus allows an even transmission of sound to be radiated across the frequency spectrum of available notes.

"Inherently present in acoustic instruments courtesy of natural physical phenomena [Hunt et al. 2002 p5], a 'mapping' layer between sound source and controller has never been necessary in a discussion of traditional instruments. Chapter five outlines some strategies adopted to overcome the essentially problematic keyboard associations (as both 12-tone equal temperament keyboards and 'qwerty' keyboards might imply) that have been regularly used in electronic music despite their overriding incommensurability with the electronic sound world itself. Andy Hunt, Marcelo M. Wanderley and Matthew Paradis make it clear that (following Rowe's 'instrument' paradigm) **"we are in the early stages of understanding the complexities of how the mapping layer affects the perception (and the playability) of an electronic instrument by its performer"** [Hunt et al. 2002 p5/6] and also note that **"by altering the mapping, even keeping the interface and sound source constant, the entire character of the instrument is changed"** [Hunt et al. 2002 p1].

However, we find that for an interactive instrument (in Rowe's 'player' paradigm) **"mapping any one line of cause-to-effect does not describe the operation of the instrument as a whole"** [Chadabe 2002 p4]. For Joel Chadabe, an instrument is interactive when **"the instrument is influenced by the performer's controls, and the performer is influenced by the instrument's output"** [Chadabe 2002 p2]. This extends to the installation setting by allowing the humans and machines present to share the supposedly compositional decisions: **"a performer shares control of the music with algorithms as virtual co-performers such that the instrument generates unpredictable information to which the performer reacts, the performer generates control information to which the instrument reacts"** [Chadabe 2002 p2]. Similarly, Stephen Wilson describes the situation thus: **"interactivity is often considered the distinguishing feature of computer-based media. The audience is invited to take action to influence the flow of events or to navigate through the data hyperspace. In the early days this relationship between the audience and work was considered quite radical. Artist and audience were seen as cocreators and the likelihood of intellectual, spiritual, and aesthetic engagement was seen as heightened"** [Wilson 2002 p653].

Now that we are no longer as tightly bound by the physics of real substances with regard to instrument and sound production, we can encourage that freedom to enter the musical world around us. Rather than say that a written score is music, or that a CD is music, we could for instance allow that music is the organization of sound in time by the listener.

Interactive installations provide an intriguing situation in which to examine our involvement in music especially as they allow one to create a non-linear view of the time dimension, to decouple the one-to-one requirements of straight-forwardly repeatable cause and effect. This process itself, though, is far from straight-forward due mainly to the nature of the technologies (the machines in their current state of evolution) that we use to play and compose music. Rather than simply extend the traditional acoustic instruments available to us either by attaching sensors (which would anyway **"interfere with the normal sound production capacities of the instrument"** [Rowe 2001 p215] and therefore disrupt or limit many normal playing techniques), or by processing the sound these instruments produce directly, we are instead looking for for a new way to articulate interesting 'places' in a vast (N-dimensional) parameter variation space of electronic music production methods that is limited only by our imagination and skill in choice and creation.

A possible solution suggested by Joel Chadabe is that one can compose a semi-automated system, **"a fly-by-wire system [that] might be viewed as a series of if-then algorithms, each algorithm triggered by a performer's action in a particular context and able to react dynamically in changing arbitrary relationships between controls"** [Chadabe 2002 p4]. This reflects the principle of balance outlined above, where some 'middle way' or 'normal path' is known around which one can measure variation by observing the world for common patterns, dealing with a statistical representation and keeping random elements under some sort of monitored boundaries. Later discussion into this way of working will therefore necessarily include requirements and techniques for altering data value and flow rates into areas to which a human being is sensitive, averaging over chunks or durations in time etc.,.

In this way, one can move from arbitrary signal representations and computational models to address the real world constraints intertwined with people and instruments producing music. Instruments here may include traditional acoustic instruments or may extend to considerations of the tools available to us in electronic situations with loudspeakers, cables, recording buffers, etc. Rather than stick to the 'note' paradigm of classical music (or MIDI), the problem refigures itself to being one akin

to musical phrasing concerns - how to group things together that 'belong together' so that multiple parameters can be influenced simultaneously according to our perceptual subjectivity and whim.

Chapters five and six discuss some basic findings in audio perception (by humans and machines) that are useful for improving techniques to control sound presentation, whether working in a stance slanted towards performing or composing. This discussion also makes clear that many of the same tools are in use in both performance and composition. Synthesis techniques have been available on paper and in theory to us for some time, but they can now be easily realized with computing techniques to bring them about, and, what is more, to run them live in real-time.

This speed of access to processed audio results is undoubtedly another factor, allowing electronic compositions to be performed live rather than being presented on a fixed-media format such as tape or CD. From an egotistical point of view, it allows a person to be more clearly the centre of attention themselves, allowing them to promote their work and more finely control the individual nuance and detail in its presentation.

Thus the gulf between acts of composition and performance is shrinking, and from a psychological stance the matter basically boils down to the preference for working in real time or not, and considers whether the creator's ego must be fed by an expectant audience or not. Whereas a performer is involved in the immediacy of presenting music live, playing the instrument, and offering a particular sound-experience to the listeners gathered together, a composer, if he is content to spend 'working' time as sole listener within the process of creating sound, does not necessarily require an audience to be present to experience the work alongside him.

The field of interactive art works blends these paradigms by containing elements of composition and performance both for the creator and for the interactor. Since instantaneous and continuous control of all things is not required as it would be for an instrument (any process may be automated to the precision desired), varied contexts of sound control and production can therefore be explored through time in a non-linear fashion by the listener.

chapter four

The previous chapters have described the complexities involved in interactive installations. With such a lot of information to keep in mind while acting, the question is frequently raised of why to hand over this power of sound control to the audience? Why let the public perform? Why let people with no experience of time-ordering order time? They haven't been trained - they don't know what they're doing! However, another point of view allows that the situation of installation art provides a platform where play is encouraged, one that allows people to contribute to their environment.

Some installation artists chose not to allow interactivity. Their work, though 'installed' and though possibly involving multi-media environments, accepts no 'input' from the public and therefore functions in a manner more like the traditional concert performance which theoretically can occur with or without the presence of an audience. Even in an installation where it has been the composer/designer's intention that the work be received in a passive, non-active, non-interactive manner, the human can of course break the rules, and behave in breach of peace or take any action open to them at any time they chose, provided no internal or external force prevents them. A point to consider is whether such an action would, like in a concert setting, interrupt someone else's experience of the work standing nearby?

The 'audience participation' factor need not, therefore, necessarily arise intrinsically alongside the installation concept, but it is interesting. Music is saying something to the audience, if it can keep its members' attention. Recalling my experiences as a 'background' musician in varied of situations from fairly formal string quartet occasions to busking on the main shopping streets, I recognize that some people want to contribute to the musical surroundings and appreciate the chance to do so - interactive works are for them.

My interest in sound (whether any one person deems it 'music' or not) has always been enhanced due to my direct participation in physical actions that result in sounds being produced, enjoying and understanding music even more through playing it than just by listening to it. Therefore, the stance that allows a much greater sense of action, or participation in the

'listener' is encouraged in my work, so that the listener is involved (more or less consciously) in the decision making process that determines the overall audio content presented. In this way, the audible sonic environment within the room can be influenced and controlled to some extent by the actions of the listener, thus giving him or her some of the responsibility normally assumed by the player of an instrument in a more traditional concert setting. For this reason, I think of the person within the installation as a player rather than as a more-or-less passive listener. The word player also alludes to the sense of gameplay, immediacy and experimentation that is frequently a perceivable feature of such an interactive audio work.

Creators of installations can also use another side-effect of this involvement of the public namely that it provides as a way to widen our world, questioning whether one is making the 'right' things accessible, whether one is guiding a musical experience or providing a method to follow. Interactivity should increase repertoire of actions and thus the chances of a good mutation arising should also increase. Are we entraining the listener in a sense similar to resonance or are we left feeling pushed and pulled around by the music, unsure where to fit it into our own continuum of previous experiences?

A balance between predictability and surprise is often stated as an ambition of such works, described as such so that with an element of safety in place due to some rules of the 'game', further rules within each particular section of the piece currently being experienced then allow specific 'moves' to be made and monitored. These 'moves' involve a process such as a live audio message from the person being translated into control data for further stages for response or the exhibited behaviour. The 'safety' of such a contained art situation can encourage experimentation and risk-taking, allowing the interactor chances to re-conceive of their role within the work. Is it possible to have a musical experience within an installation? Does interactivity matter?

Allowing the general public to take an active role increases, for instance, the chances of surprising discoveries being made, such as when someone approaches an unfamiliar context and brings about audio results that had

not previously been conceived of or heard by the installation's creator(s). For example, during the presentation of "So They Say", an interactive audio installation presented by myself and Matt Wilkens (at Come to Ur Senses festival, May 2004), a video editor who participated discovered a method of vocalizing particular sounds over long periods of time so that particular layers of quasi-periodic rhythmic patterns could be created. His skills at working in segmented chunks of time and applying batch manipulations in image allowed him to recognize a particular feature in the audio result, and thereby he learned to feed it with sound at specific points to expose a long-term emergent behaviour of the installation in audio of his own devising.

The situation becomes still more complicated when we consider a multi-person interactive environment. Is the will of an individual in alignment with that of the rest of the group? Does a performed action have such a result so that a subjectively 'good' sensation can be continued, extended or heightened in some way rather than decreased? Can the intentions of the player be transmitted in such a way that he or she can feel some freedom in self-expression? Chapters four and five will describe what sorts of 'expressive data' can be gathered from the environment surrounding the player, whether by audio or by other means of registration. This chapter, however, first addresses questions about the nature of these interactions in art that share decision making processes between collaborators, reconfiguring the relationship between the so-called audience and stage.

Very many questions still surround the purpose of music, particularly in an interactive installation, and the form of participation must be planned carefully. Is there an element of guidance given to the participant? Should it be popular or alienating? We must ask - what informs our actions as interactors? How do we judge an interactive experience? How do the cognitive and emotional and behavioural aspects balance out? That is, how do we judge the performance in action? Do we enjoy ourselves? Do we benefit from the experience? Do we breed an increased openness to interactivity in our children?

Di Scipio discusses the implicit feedback loop present in most interactive system designs [Di Scipio 2003 p270]. Although on first glance many so-

called interactive music systems seem to suggest a linear flow of information from performer to the output sound (through control devices and the computer signal-representations), a recursive element is actually implied such that the human's listening and judgment processes of this output sound may in turn affect his or her subsequent actions (which in turn would alter the following signal chains again in the computer, and so on and so forth). Many works in this vein rely on concepts such as action and situation to describe these sorts of processes on different time-scales. Robert Rowe describes "situations" as **"represent[ing] relatively static states of the environment"** and "actions" as **"always [havng] an initial situation, from which it is begun, and a final situation that it produces"** [Rowe 2001 p374].

When an installation artist is then working to compose the interactions that should bring about the desired environmental conditions in the space, it becomes necessary to define tasks such that 'the necessary events happen' - whether sharing these tasks simply between creator and the public, or more democratically between different artists working together in collaboration, the processes are similar. Interacting suggests that one must accept the responsibility that some freedoms bring and yet also accept that there are restrictions and non-freedoms in place. One must in every case attempt to align ones own power to the regions where ones efforts will produce the effects one desires, searching for the right action to make at each moment, subjectively feeling the way at uncertain times to seek out the interesting points and find ways to articulate them. Actions can happen self-consciously, not just with self-obsessive internalized reflection but actually in judging the effects of one's actions in the world to determine whether progress or regression has occurred.

In their design of an interactive environment, an artist can place a great responsibility on the interactor, altering the 'normal' relationship of viewer and art-work. Such an installation can be crying out for a player to come along and make the music happen, rather than using its visitors as pure observers to the music - the traditional role of the 'listener' is thus vastly extended by the introduction of machines that are sensitive to particular portions of the atmosphere of its context in the environment. Really, the crucial aspect of such a work is that it has shifted the stance from the importance of music in the player to one of considering the

importance of music in the listener. Akin to what occurred in Cage's 4'33", the interactive installation now allows for **"shifting the production of music from the site of utterance to that of audition"** [Kahn 1999 p158].

It is helpful also to question cases that lie outside of the ideal easy interaction environment, for example, that of the reluctant participant. Further, is there such a thing as a non-interactive person? If so, would they have the same feelings as interactive people? Shyness, or nerves for instance, may overcome a participant (or indeed they may suffer from a more serious, even disabling condition) so that their desire is not to 'join in' but rather to retire to seclusion, feeling that the state of the world is such that the prevailing conditions prevent his or her execution of a successful action. Robert Rowe reminds us that **"the computer is a better member of an ensemble when it has a sense of what the group as a whole is doing"** [Rowe 2001 p312] - perhaps this comment can be applied equally to people?

Traditionally, music is always of bigger scale than one individual person - personal choices allow that one may chose a small area to work in, then to practise this - the whole field is thus narrowed and ordered. Performers carry out this act by selecting their instrument, their period (and often geographical location) in history, and then set about learning the examples of works contained within this set. Interactive works, however, generally aim to be 'open' so that a person can approach and chose a level on which to deal with the sound around them, each according to their own talents. Ideally, interfaces try to open themselves to capture something that will, in the end, increase the musical knowledge and skill of the individual using it and simultaneously that further the progress of the music around them. In some works the audience's role is reconfigured like a conductor's - that of directing an otherwise unorganized group of potentially loud sounding instruments - building an environment, a safe one, where one can practice the close and direct control over any particular manipulation of parameters without causing the next world war, without causing the sky to cave in around you or swallow you up!

To deal with the situation of different types of people who may enter such an interactive space, especially when multiple participants can be accommodated together, it is common to see an installation with multiple

levels of input 'control' that result in musical output. The premise is that interaction can help the visitor to situate themselves more quickly in the moment of 'now', finding their place in the present and just 'being' in sound, seeking the changes that one needs to perform to bring about or pull the current situation, if possible, into line with what one would like the current experience to be of.

To increase the chances of the experience being interpreted as 'musical' in such an installation, even for non-musicians, some installation artists chose to create an environment that will guide one's actions so that a participant will gradually learn to perform actions requiring musical skill to complete. One such example is seen in "The Interactive Dance Club" [Ulyate and Bianciardi 2002], a multi-participant work which, in the confines of a club rather than concert hall or gallery space, allow several simultaneous zones for different types of audio and visual interactions to occur. Another line of inspiration chosen by some installation artists focusses more closely on the sense of play inherent in music, aiming to encourage that "purposeful purposelessness" referred to by Cage and elaborated further as a type of playing with music that is **"in turn explorative and engaging, intuitive and enjoyable"** [Robson 2002 p50].

Perhaps anticipating the fact that attention spans are short and the pressures of modern living do not allow much free time to want to spend long attending an installation or other work of sonic art, other installation artists situate themselves more in the tradition of Satie, whose output is variously referred to as 'furniture music' and 'wallpaper music' - **"music that would be a part of the surrounding noises and that would take them into account"** [Cox and Warner 2004 p65]. These comments refer to his pieces, often just a few bars long, that be would repeated over and over ad nauseam, not so as to impinge far into the consciousness of the listener but rather so as to provide an atmosphere or aural back-drop to a particular location or situation. Some try rather to block out the noises natural to the environment, and use tricks akin to those of 'dramatic music' in theatre in the seventeenth and eighteenth centuries, with large dynamic level changes and so on to draw attention to the work itself.

There is a type of listening that uses music to create a space and time for

us to be quiet in, to be reflective, to look inwards, to be on our own. The concert hall provides for this (though in a strange balance of public and privateness that does not suit everyone - some report feeling 'stuck' inside these halls, others fidget endlessly whilst distracting everyone close-by). A gallery often provides for this type of listening (hushed talking in small groups is considered acceptable behaviour whereas shouting across the room usually is not) and, often even the club environment provides a space for personal isolation (as the music is too loud to bother trying to communicate verbally with people around you, so small would be the chance of a particular message being relayed).

As the contexts in which we situate installations and especially interactive art works widens, the study of the aesthetic questions involved must gradually broaden to allow a clearer understanding of what we face when an audience (of one or of many) is included as an aspect of the work itself. Like introducing immigrants to a foreign country, are we dealing with concepts of integration or assimilation? Are we fighting art's separation from 'real life', reassessing art's place by moving from the paradigm of performed 'events' with specified conventions of duration instead to performable 'areas' open for non-expert exploration in sound? Is the biggest resistance to interactive works likely to come from an audience steeped in western tradition and looking for personal quiet-time for reflection in otherwise silent listening conditions? When an audience, en masse, is hungry for a particular type of sound and has the power to bring it about, what happens next?

Music has a power over people, and there are important factors of health to bear in mind. Quite apart from the torture methods and ways of inflicting physical damage using sound that have been developed over the years, and apart from the historical uses of instruments in warfare (where the bagpipes, it is said, would perform the double function of drawing the troops together and of terrifying the enemy), we now find music being used as a way of entraining people into a particular mindset and even as a way to reduce the intake of pain-relieving drugs during operations in hospitals. How much power should an untrained listener be given? In order to maintain a safe and aesthetic environment one must attempt to solve **"the problem with ensemble improvisation ... [that] is one of designing an appropriate control structure"**

[Rowe 2001 p310].

"Triggering Daily" (April 2004, Bremen) is a work inspired by many such questions. Whilst working on this collaborative project with Claudia Kapp we referred to 'an installation to be performed' comprising an environment with six microphones at input. Four of these were dedicated to individuals, but the other two were open for contribution from the public - one microphone was dedicated to the seated audience in the immediate environment hanging just above their heads in the centre, and the other microphone hung above the pavement on the street outside the door to the venue (visible through the glass-front of the building). Playing with ideas of imitation, transmission and infirm boundaries (specifically with relation to both humans and machines), the work uses an internal repetition and fading-delay system to build four textures of sound from the recent inputs of all six microphones. However, rather than play these sound layers into the room directly, these layers are fed back to four sets of headphones, one layer in each. Four vocalists are involved, each listening to a layer of sound and imitating what they hear (each differently) directly back into one of the four microphones reserved for the individual performers, and it is this output sound (four human copies of computerized audio sound) that is played through loudspeakers to the audience seated in the room. Thus, the set-up categorises quite traditional electronic processes and materializes those concepts on a very physical and personal level (literally, one person per 'memory bank' or delay-line texture).

The main concept underlying the development of "Triggering Daily" was to find a way for the audience to become a part of the show, to make the members of the seated audience hear that their presence was an important part of the events. Therefore the work began with all four vocalists attempting to imitate the noises coming from the microphone situated just above the seated audience - small shuffles, coughs, whispers and stifled sniggers - before introducing the elements of outside (the rhythms of falling footsteps, snatches of voices from passers by, a dog barking, cycles, buses and trams) and then chaining processes where the sounds from the vocalists themselves also enter the delay lines in the computer. Perhaps the most instructive moment in the work for the seated audience inside happened at a point where quite a crowd had, by chance, gathered

outside. Those outside, able to see what was happening but not to hear it, gradually came to be making more and more noise until a young girl was screaming at her father to let her go inside. Much of this aural action had been percolating through the various textural delay lines into the headphones with the associated imitated versions then being reproduced vocally in the room. When this group of children finally did enter the room, the seated audience cheered loudly, a sound which gave another texture entirely new through the headphones and re-transmitted through the various vocal imitative styles back into the environment.

This project also highlights the question over whether as listeners we chose to sit down and let the sounds wash over us, or whether we try to engage with the sound directly. The audience, seated or outside, were involved consciously and unconsciously respectively. Similarly, each of the four vocalists had the same 'task' - to reproduce what they were hearing - yet each interpreted similar computer-results to produce very different vocal sounds, each steeped in their own prior experiences (singers, sound artists, spoken word artists, etc., each individual's history playing an audible part in the final performance). The plurality of different hearing methods was thus exposed, a fact which will be discussed again in Chapter six.

The "Triggering Daily" project is another example of a collaboratively-devised project, reminding one once again that **"artistic flexibility and the willingness to pursue nonstandard paths are resources in this work"** [Wilson 2002 p653]. Many things in music must be done without a literal definition of what is meant - discussing tricky concepts forces us to learn how to communicate about the tools we have available and the results that can be produced with their reconfiguration in whatever manner can be imagined. Being an electronic musician, especially when working with people whose first language is not one's own, requires that one can adapt and bend one's individual efforts to fit the 'greater good' of the combined effort. Just as the interaction between machine parts must be defined within the realms of its action, so too must the roles that each person plays within the collaboration in order that they can continue to be focussed on the final 'product' that is the finished work. Therefore it is important to see how the areas of work are to be divided, to monitor how work in each of these things is progressing,

and to look out for the potential pit-falls beyond the immediate situation before they arise.

Collaborators, often working in different media, have a chance to evolve a means of communicating about certain themes, illusions or real patterns, and in their meetings (whether peaceful or more like intrusions piled one atop another) can together establish a framework that allows the individual elements present within it to exist in certain relationships with one another. Written texts are generally held to allow a transfer of ideas across times and places, but each reader brings their own associations to the words before their eyes. The act of collaboration usually involves more spent time in a face-to-face working environment, programming or listening to sound material, and we must therefore learn not only to understand one another in word but also in technique. It can in fact be far easier for people from different backgrounds to communicate when there is a shared environment for working in (for example, the MaxMsp platform) - though each programmer has his or her own style of work, time spent sharing and understanding the patches programmed can help express conceptual ideas in much the same way that a violinist may, with exaggeration, sing a phrase to a student to suggest where changes should be made.

With respect to getting machines to do what we want them to do, each programming environment introduces a very limited but shared vocabulary by convention into that area of work. Whereas in Western art music, the discussion is of tonal systems and harmonic progressions and so on, electronic musicians talk instead about frequency values and fft window sizes and so on. Many mathematical terms are currently saturating the discussion of electronic music and installation art. Human-computer interaction and computing terminology abound also, and yet each artist, individually motivated, reconfigures these elements with previously undreamed-of artistic freedom. We may wonder how electronic musicians can share body language in common with each other, and further, with players of acoustic instruments? The following chapters will begin to address some of these issues in the hope that one day there can be a far closer and more immediate relationship directly between humans and machines so that discussion can return once more to the musical intentions of behind each phrase played rather than constantly being confronted, as one is now, with

questions of which parameters should be kept control of now, and when should they change next.

Interactivity, loosely, is about sharing ideas around - with each other, whether that 'other' be a collaborator, a visitor to an installation, or indeed a paying customer. Collaborations also have another very obvious advantage - that of a shared workload. Without dwelling on the topic, it is worth noting that promotion, marketing, and further commercial aspects to do with funding research and paying for equipment, space and electricity are all easier to take care of with the support of colleagues. The commercial nature of the world today, a society that is increasingly likely to sue for damages at the slightest provocation, suggests that it can only help to have collaborators with whom to share any extra layers of paperwork. Intellectual copyright implications are infringing into many areas of music, and musicians seem unable at present to agree wholly over concepts of theft since it no longer has the connotation of removing an item from someone else's possession. Having arisen through compositional techniques such as sampling, such concerns already affect the music that is made, and leaves one wondering how this situation may develop in the future. A publication in the '90s made it clear that **"the pressure for keeping any recognisable form of public domain has become entangled to an unprecedented degree with questions of copyright, broadening the whole issue of changes in access to music from specific questions of the cost and availability of machines to issues of law and cultural policy"** [Durant 1900 p187-188].

Alan Durant continues with another point of concern, **"Even apart from deliberate copying, there has always been the practical problem that, since individual creativity in music necessarily involves innovating within established genres and styles, the risk exists of repetition, imitation, and thereby, in this century, of copyright infringement"** [Durant 1900 p191]. Even should one manage, as artist, to keep one's work entirely original in material and infringement-free, perhaps one's collaborators actions should be considered also (whether colleagues or interactors). This raises an interesting point specific to interactive audio installation works where one allows a portion of the live input sound to be reflected back in the final audio: what happens if an interactor choses to play copyrighted material into the microphone input to the system? Is it implied (by unspoken rule) that a sound artist ought to devise a way to remove such material from the system? This, like many other areas of modern life, is an

aspect that is developing so quickly in recent years that legislation is slow to catch up.

chapter five

A truly successful interactive audio environment can be said to allow an exploration of time and space in sound. Time and space are, after all, both required for people to perceive audible sound. The space allows an interval between events in time. The interactive experience (that is, of a person existing in time and space) can then be explored through spatial and temporal movements, seeking correlations in the relationships between the physical, bodily, real contexts of the work, and the perceptual experiences arising from the resulting sound. A view commonly accepted in social and scientific practices states that **"our reality is shaped by the patterns of our bodily movement, the contours of our spatial and temporal orientation, and the forms of our interaction with objects"** [Straus 2004 p15].

The focus for us will return, in the following chapters, to the various audio representational methods available that increase the success and efficiency of computer listening methods, but it is helpful first to consider some related work. Analysis, whether of audio or other data, is essentially about opening a window of opportunity and then receiving and interpreting any signals that exist within that window. Thus, fundamentally, our concern is over the content and timing of the flow of information. Video detection methods, as employed by the Kansei group for instance, have allowed the 'meetings' of humans and machines to receive a very considerable amount of attention, and suggest that there can be an 'emotional' significance attached with our actions [Camurri et al. 2002].

Rather than dealing in pure audio signal representation, this chapter examines the step between a player's physical action in the world and the possible manipulations of audio content resulting from this action. There is no clear, unambiguous way to define how a movement should be used to modulate the audio environment, nor how any measurable parameter may relate to a musical idea, but as I believe this to be a rich area of possibility open for exploration I will outline some findings here.

Instrumental teachers concentrate much of their effort on how the body is used in order to manipulate or control the instrument, to 'make the music happen'. They talk about how to handle the human (body and mind) in order that the instrument may be excited to produce the sounds held in the

musical imagination. The player then proceeds by continually considering what they are actually hearing, that is, the audible results of their own actions, and comparing this back with the concept held in the immediate consciousness of the imagination about how their own sound comes across. Whether playing alone or in a group, this continual assessment should be happening inside the player's mind alongside any actions taken physically.

Briefly revisiting the ideas presented with regard to the differences in composition and performance, we can analyze these acts with regard to the metaphor of reduced subset classes of choice, even stretching the point to a large-scale implementation - instrumentation in the symphony orchestra. A viola player cannot ordinarily decide to suddenly change instruments and play the clarinet or horn even in the case that those instruments share a musical line (as is often the case in music by Brahms). Though the instrumentalist may perhaps be physically capable of such a thing, the instrumentation has been fixed in advance and the conventions of the performance situation forbid it. The composer has made these decisions ahead of the time of performance, usually by creating a written score or textual record where some information is preserved that allows a more-or-less faithful recreation of the original idea.

While these decisions may be informed by the instrumental qualities (register, timbre, articulation possibilities, etc.), the composer does not overly concern him or herself with the actual physical movements a player must make beyond questioning whether the part is playable by a reasonably competent performer within the technical specifications of the selected instrument. The player, however, to perform a wide repertoire of music, must train solidly and place trust in their muscle-memory so that groups of actions can be performed without having to consciously build the larger 'movement' out of its smaller blocks of action. Alongside the speed and stamina requirements of great virtuosity in many works, a player will often exhibit slow time periodicity in body movement. Feeling this slow pulse can direct the listener's attention through the phrases, can make one anticipate the resolutions and potentially enjoy the experienced tensions in their gradual release rather than simply 'clocking' every individual beat or moment that passes.

The situation must be reviewed considerably for the case of playing electronic music. In the past, issues arising from the newness of developing technologies impinged primarily on the secondary aspects of the music business - in printing, recording, instrument manufacture and MIDI networking provide good examples. However, as technologies continue to develop further we are now finding factors arising in the primary aspects of music with respect to the use of technology - in listening to, composing and playing music itself.

Perhaps the biggest problem, and simultaneously the most rewarding work, facing an electronic musician today is that one must find a new language in which to articulate music. Whether steeped in the classical system of western art music or whether driven by the groove of expressive hand actions on a drumhead, one must now learn to communicate with machines. Computers understand the world in terms of ones and zeros represented internally on differing scales and resolutions. Mathematical concepts and all manner of trickery with soft- and hard-ware exist or can be fashioned, and must next be configured in order to reduce our cognitive strain to the point where one is able to simply play music again.

For our ancestors, the music available was generally restricted by one's geographical location although now this bias of cultural exposure cracks in turn since music, instruments, software and ideas are increasingly shared internationally around the globe. We are situated now in a land where opening oneself to a cultural moment is a choice selected under more aesthetic conditions than would ever have been possible in a world without aeroplanes and the internet. This idea, like many that are central to the choices that must be made in music produced today, is exploited in science fiction. One parallel aching to be drawn is with Neal Stephenson's book "The Diamond Age, or A Young Lady's Illustrated Primer". The action surrounds a street-girl who, through the use of an interactive book, is able to overcome her humble beginnings and learn of other cultures. Born (relatively) neutral in this land, sooner or later one will join a cultural group or tribe, perhaps aspiring to live according to Neo-Victorian mores, and will thereby accept a set of more- or less- typical conditioning behaviours, an attitude by which to interpret the contextual information of living.

By analogy, we now must question and select which methods are most appropriate for us making music in the electronic age. The emphasis on storing a linear text which unfolds across time has lessened in electronic music - a far greater number of working hours are devoted instead to synthesis methods and how to control the sound bodily, physically, in ways that relate to our listening capabilities as humans. We are at a paradoxical point where we would like to produce an archiveable record of an event at the same time as producing the archiveable event itself - yet installation works often take place in a living museum or gallery, and come into being without any easily transcribable form that could substitute for a written score. Without the written record (the token that culture would retain) we find a far greater emphasis on experimentation in this music, channelling many different cultures into an electronic medium and exploiting the fact that, to a computer, any arbitrary mapping of real world input data is equally valid.

A performer's movements are of central importance when one works with dancers and, as a result, a considerable amount of research has been done to classify particular movements and body positions in order to use these as a compositional element in new media stage and installation works. One approach is to attempt to teach the computer something of how the traditional audience must 'feel' when seated facing a stage. In this case, when ready to 'hear' or 'see' or 'experience' something, a mediated signal that draws on a particular aspect of the performance can be absorbed into the computer from the stage, and interpreted according to an internal template or scheme of representation that might allow some 'understanding' to occur.

The Kansei group approach this topic aiming to enhance the interaction between man and machine [Camurri et al. 2002]. Three levels of processing occur in their model: firstly relating to the physical signal itself, secondly on linguistic and semantic aspects of the situation, and thirdly on the Kansei signal which involves the realm of feelings, intuition, and sympathy. Software developed by this group under the name of EyesWeb interprets the data it receives as coming from two areas - a 'general space', and a 'kinesphere' representing the personal space around a body.

Their work then tries to find places of musical relevance which relate to the movement data from their analysis (silhouette shapes, quantity of motion, body contraction indexes, motion segmentation and gesture representation). Similarly, movement data also has the potential to be extracted from video signals by the EyeCon system developed by Palindrome in Germany, by David Rokeby's VNS (very nervous system) in the US and with the BigEye software developed at STEIM in Amsterdam by Tom Demeyer.

The situation becomes more complicated in an environment where several performers are involved simultaneously. To look again within the classical western art world for an example, we can consider a string section in an orchestra. Socially, the players conform to an hierarchical organization, structuring the people in such a way that they 'play together' and thereby present an integrated 'overall' picture to the listener. While a well-trained viola section will, of course, 'sound' together, visual clues such as body-sway or bow-direction can help to integrate this movement as one source ('the violas') in the listener.

Levitin, McAdams and Adams [2002] see such motions of the body as gestures that contain information. Their paper describes considerations to keep in mind when mapping a musician's gesture as a controller for sound, and suggest using intuitive translations that make use of our inbuilt cognitive map and physical properties of our planet. Thus 'hard' suggests loud, 'tight' suggests high, and a 'wiggle' could suggest vibrato.

This outlook is supported by concepts of embodiment in fields of philosophy and linguistics, more particularly by a theory sometimes called 'experientialism' which music theorists have now applied to musical hearing, stating that **"the way we listen to music, hear music, understand music, is shaped by who and what we are"** [Straus 2004 p14].

In the role of 'interactor' therefore, this implies that each experience is contingent upon one's own previous experiences, dependent upon the "cultural baggage" brought into the room when one enters [Daubner 1999]. Perhaps conscious of this, the installation artist David Rokeby believes that **"an interactive technology is a medium through which we communicate with ourselves - a mirror"** [Rokeby 1995 p133]. He continues, **"a technology is interactive to the degree that it reflects**

the consequences of our actions back at us" [Rokeby 1995 p133].

We must therefore turn to what may be termed 'significant' with regard to sound and body or mind. Realizing that **"a morphology imposed upon electronic sound-objects through the monitoring of performance gesture can be much more refined and subtle than that resulting from intellectual decisions made in the studio. The directness of physiological-intellectual gestural behaviour carries with it 'unspoken' knowledge of morphological subtlety which a more distanced intellectual approach may not be aware of"** [Wishart 2002 p179], the task for the composer of interactive situations then becomes one finding **"physical handles on phantom models [that] stimulate the imagination and enable the elaboration of the model using spatial and physical metaphors"** [Ryan 1991].

Many people are currently researching different 'parameter mapping' models that seek to achieve some perceptual correlation between the computational variables in use and the sound results as experienced, using mappings that encourage the player to think about sound in perceptually relevant, and perhaps, even in traditionally musical ways.

In searching for better metaphors for live performance, the CNMAT team offer some basic hints, for example that the size of gesture should be proportional to the size of the perceived result [Wessel et al 2002]. They continue by outlining spatial clues that can be exploited in time (proximity, drag and drop, scrubbing, dipping, catch and throw, and listen and respond) in order to further elaborate a musical dialogue. These metaphors are selected because they have an underlying basis in our perception, making use of **"both the way we are structured physically as embodied organisms and the structure of our environments that permits only certain kinds of interactions with those environments"** [Straus 2004 p17].

We have now encountered the benefits that such metaphors can bring to performances using new technologies, and so we find now that **"gesture may be seen as another musical parameter to play with in the context of interactive musical performances"** [Sapir 2002 p119]. However, care must be taken to heed Joel Chadabe's warning and remember the limitations of mapping systems when applied to interactive works where the control of audio is to be shared between people and machines, that a one-to-one cause and effect mapping is not sufficient. Di Scipio highlights this fact in his criticism of many so-called interactive systems where **"the computer's internal state depends on the performer's action ...**

the latter may itself be influenced by the computer output. One could introduce complex transfer functions in the mapping of information from one domain to another ... but that is not essential to the underlying ontology: [human] agent acts, computer re-acts. ... Indeed, in this common notion of interaction, the [human] agent is indeed the interface between the computer and the environment, and, at the same time, it is the only source of energy and change" [Di Scipio 2003 p270].

Rather than fall into the instrumental trap, where the system of sound-generation is unable to change its own state without a human performer being present to direct the change, instead, the computer can expand the musical field and enter into an exchange with the player, perhaps by creating surprises within a controlled framework. One can, for instance, explore in audio the relationship between real and delayed time processing. Alternatively, one can express some relationship between the sound signal and the symbolic musical structures created in the artistic presentation.

Michael Hamman states that **"an interaction is an exchange of energy between two different existents, be they imagined or physically realized. An "interface" defines the mechanisms - physical, conceptual, and cultural - by which a set of interactions appropriate to a particular cognitive domain is engendered. An interface can be as simple as a door knob or as complex as an airplane cockpit, or it can delineate a conceptual framework such as a text or a score"** [Hamman 1999 p91]. Generally, musicians have highly trained hand-eye coordination skills that make them useful in certain obscure ways. Some controllers allow a portion of this muscle-memory type of musicality to be transmitted, and attempt to address the dimensions crucial for music by **"determining [and re-applying] the rates, directions and probabilities of a flow of a quantity called information"** [Harraway 2000 p302]. From this we can see the importance of matching the type of interface to the type of interaction required. Bearing in mind that **"the conventional physical computer interface of keyboard and mouse come with significant cultural baggage, ... researchers and artists read human actions such as motion, gesture, touch, gaze, speech, and interactions with physical objects [however] even these body-sensing technologies ... carry their own histories - for example, in military research ... [and] can easily be perverted and assimilated into narratives of control and surveillance"** [Wilson 2002 p729].

Notions of 'play' have been seen to be highly dependent on both the particular technologized tools selected for the tasks of sound production and the manner in which these tools can be utilized and altered by participation from a human. Rather than follow the narrow and unhelpful concept of interactivity, missing the point that new musical structuring

options in time may be constructed, as given by Todd Winkler (who states that **"a musician is free to play any kind of music, and the software has enough "intelligence" to respond in a way that makes sense and naturally elicits the performer's continuation. Like a good conversation, interactive music succeeds when it encourages spontaneity while residing within the boundaries of a dynamic musical context that is whole and engaging"** [Winkler 1995 p1]), it is more instructive to take another example along the lines of Di Scipio's that was discussed earlier. With regard to interaction in interactive art, we can apply Norbert Wiener's words - the secret of its organized action lies **"in the intercommunication of its members"** [Wiener 1961 p156]. In this way we can perhaps begin to conceive of dynamic art objects unconstrained by the standard linear stepwise motion along a timeline. Time need no longer run from left to right, from start to end.

By concentrating on audio analysis in the remainder of this thesis I am leaving detailed work on bodily derived performance data to other people to undertake, but a brief overview has been given here in order to acclimatize the reader to such practices. Many of the findings related here, due to the cross-modal nature of installation art, can be generalized from one domain to another and thus have relevance directly within the audio analysis techniques yet to be discussed. Many people continue to work on the man-machine communication aspects, modeling either physical gestures or musical structures: the implication on real-time improvisation allows more aspects to open up to 'present-moment' type of experimentation. Following from this it would be nice, one day, to see an interaction of the musician and the model of experimentation itself.

chapter six

Previous chapters have outlined the problems faced by the installation artist who seeks to involve an active public in an interactive work. Input devices (qwerty keyboards etc.) introduce strict limitations in what can be achieved, and have specific non-musical associations in the minds of the interactor. Rather than spend energy in a search for what may be more musical about one action over any other I have concentrated work on audio signal analysis and re-synthesis, developing some listening units that attempt to derive information from the actual sound present in the environment and from that to deduce aspects of human behaviour.

Interactive works can question how people relate both to individual sounds and sound structures in their immediate environment. Bearing in mind the sorts of resolution and sensitivity conditions previously outlined, and finding also that the conceptual framework for academic discussion of interactivity in music is as yet fairly loose and undefined, it is not the aim at this stage to be explicit about specific perceptual relationships.

Rather, a working method is described that draws on the considered limits of psychological impression in regard to the specification of the detail required for audio and interaction models. This explores concepts of interactivity, potentially giving more responsibility to machines to direct musical situations than has been the case with the traditional instrument metaphor. The aim is to let the machine, like the well-behaved human, listen and think before any action is made.

Whatever compositional decisions are taken, whatever intellectual attitude the composer adopts, we find that **"at the level of [human] perception ... archetypes of natural morphology and the interpretation of human gestural actions will enter into the listener's interpretation of sound morphology"** [Wishart 2002 p189]. Wishart continues, **"different kinds of intrinsic morphology affect us differently and this is something to do with the assumed physicality of the source (which is not the same thing as source-recognition)"** [Wishart 2002 p181]. This is clearly highlighted in a research paper by Daniel Freed [Freed 1990] who investigated a single timbral attribute of sound, which he named the "perceived mallet hardness", of hits with different beaters on various 'instruments' (cooking pans). The results of his experiment showed that the listeners were able to judge the PMH consistently, regardless of

which pan was being hit with the mallet in question.

Before returning more fully to the topic of human perception, we must note that when computer perception is questioned through audio we have only a single dimension to work with as input - that of amplitude values in time, representing the pressure fluctuation around mean air pressure as transferred from the microphone at the sample rate chosen and according to the amplification present. Investigations are quite different in sound when we consider whether the listener uses ears or microphones to hear, but they have in common a point noted by Cage, namely that **"structure [is] determined by duration, which sound and silence share"** [Kahn 1999 p181].

In order to better guide our computational listening models to find the 'correct' (perceived) information from an audio signal representation, we need to be able to define what we are seeking. It can therefore be helpful to ask what a human listener hears in the sounds presented to us. This, however, is a non-trivial task. Remembering that our realities are moulded by the conceptual structures and frameworks we use to model that reality, we find that different listeners hear different structures in sound, and indeed that a single listener can be more or less aware of particular aspects depending on the context of their listening situation. By analogy, to hear different aspects in an audio signal the computer must also be given different contexts within which to listen.

Wishing to give the computer optimal chances to understand humans and our musical intentions, we might consider which aspects of sound are heard by musicians (who after all have had much training in using their ears). Looking to my own, somewhat reluctant, experiences performing on various instruments over the last twenty-four years and as many years considering the nature of sound I still cannot decide what the most important aspects of sound are, even to myself, unless I tie in these considerations to a particular musical context. When playing violin or viola in an orchestral setting one is trained to blend one's sound with one's section (thus making it 'indistinguishable' in the texture). When performing thus, or indeed on piano, one is trained to listen ahead of where one plays and to judge actions to be taken on the timing grid of what is just about to come. When playing the Ugandan xylophone or percussion instruments in a Brazilian

Samba band although the mental considerations and aspects of sound to listen out for are quite different, the same attitude is present in putting over, ideally to the best of one's ability, the music itself.

When not involved consciously with the performative aspects of playing, but simply listening, we find that this process of listening has already been conditioned by our past and current experiences. For example, when hearing an orchestral piece the viola line will often come to the fore in my attention - particularly if I have had the experience of performing that work and can imagine the physical actions that bring about each sound. When composing or imagining sound or music, one can hear it internally, in the mind's ear. However, the interactive installation art world has neither a long-standing repertoire nor such contextual clues to draw on in order to present such an integrated impression to the human. The exploration must begin anew, searching to discover how music can divert one's attention to different things, playing on the fact that, say, drummers hear rhythms first, singers hear melodies, bass players hear bass lines, and viola players hear textual fillings. To many people, even trained musicians, it appears that electronic music composers hear noise.

Research methods into human hearing often result in auditory perception models. These are situated somewhere between the two extremes of a model based on the 'cause' of the hearing (such as psychologists use, considering stimuli and biological response) and one based on the 'effect' of the hearing (such as engineers use, factoring in for known limitations such as masking etc) [Robinson 2002 p56]. However, we must take care in our cultural assumptions and not rely solely on musically trained subjects as **"the situation is that novices do not resonate to octave similarity; they often cannot identify intervals as members of overlearned categories; they seem not to know online what chromas they are hearing; in many situations, they may even lack abstract chroma and pitch classes; they seem not to appreciate that the different notes of their scale have different functional and closural properties; they have little desire for syntactic deviation or atypicality; they dislike the formal composition elegance that characterizes much of Western music during its common practice period; indeed, they reject music by many of the composers that experts value most"** [Smith 1997 pp 251-252].

Scheirer reminds us, **"[this is not justification to] dismiss non-expert listeners as musically "worthless"; rather, it is to say that unless we want music systems and theories of music perception to have relevance only to the skills and abilities of listeners who are graduate-level musicians, we must be**

cautious about the assumptions we follow. Naïve listeners can extract a great deal of information from musical sounds ... [They can] not only extract, but preserve in long-term memory and reproduce vocally, the absolute tempo and absolute pitch of popular songs that they like—and they are likely to be the primary users of many of the applications we would like to build" [Scheirer 2000 p37].

The electronic sound world is a very strange place for the average person: electronic ears are deemed peculiar and obsessive, and the relationships outlined in structures of the music are often out of reach of many people's attention, let alone their definable consciousness and understanding. However, by allowing an individual to have some 'say' in what happens in the audio presented live around them, allowing them a chance to experiment with(in) sound in a safe environment, enlisting their 'help' by encouraging their participation, a small contribution can be made towards opening up more minds to this particular world of sound.

Though it may sound somewhat tautological to mention, it is worth repeating that the primary concern in an interactive audio installation should be audio. Previous chapters have begun to hint that **"no objects, spaces or bodies are sacred in themselves; any component can be interfaced with an other if the proper standard, the proper code, can be constructed for processing signals in common language"** [Harraway 2000 p302]. We wish to instigate a situation where, alongside the musical output (in terms of audio signals sent to loudspeakers) we also allow that musical input (in terms of audio content presented to the microphone) may be the device or controlling mechanism for the environmental conditions.

Further work in this area allows more evolved interactions to be achieved with sonic means, and thus a deeper understanding of interactivity develops through using sound itself as the interface between humans and machines, and between their combined and reconfigured presences in environmental (or to use Di Scipio's word, ecosystemic) spaces. While synthesis and performance environments have been well documented in fields of art, listening methods have often been relegated to scientific research projects and more recently voice-recognition control tasks, but have rarely been exploited in sound.

One task is therefore to build better or more musically-useful listening models for machines, and a second is to work on how these can be used in a musical context. For the performer and composer alike, an eventual benefit

may be that electronic musicians would more easily and exactly be able to compare their sounds with one another on stage or in the studio and thereby potentially improve the quality of the music that they make, spending more effort in musical decisions rather than in technical work. For the sound installation artist, the limitation of working purely in the audio domain allows a tight focus on detailed working methods and allows one to attempt the creation of environments that question aspects of musicality (or otherwise) in the participant. Today we are almost saturated with music: in shops, cafés, bars and in fact in nearly all public spaces. It is so common to hear music that we often actually tune out and switch off, not hearing it consciously. It is hard, at least in the so-called developed nations, to find someone who has not been exposed to such a constant flood of music. Rather than follow this path of ubiquitous music production, by providing ways to encourage participation and allow people to once again take some control of their audio surroundings through specifically sonic means we can allow both the trained-musician and non-trained-listener some means for engagement and various paths towards satisfaction.

There are at least two different (complementary) ways that a machine can deal with an audio signal as input: either processing audio signals directly to produce transformed aural sound layers, or by using the audio signals instead to derive specific items of control information. This latter use of the audio input is the key issue within research today which aims to explore the sorts of things that claim to be 'dimensions' present in music: feature extraction from complex audio layers.

Previously, listening models have concentrated primarily on the frequency and time domain approaches to analysis, allowing some audio tools such as pitch-trackers and beat-trackers to be developed. As a simple model for discussion, we can think about the frequency- and pitch- dimensional correlation. This proportionality, however, holds only within certain frequency- and pitch- constraints, and does not define the other factors (such as loudness and timbre) that can also have a part to play in the perceived pitch. Therefore careful consideration must be made of the limits of the perceptual or physical model of the mapping system used to get 'somehow' from the human input data to the interactive process. The sensitivity (that is, resolution, or scale of differentiation) along any

such perceptual dimensional axis is therefore another implied factor to bear in mind, such as is seen of the division of the octave into say twelve equally tempered semitones. A musical transposition in pitch, then, may have defined relationships existing in only certain regions for a specific and perceptually meaningful alteration of the pure frequency information in the audio system.

Commonly discussed musical concepts such as pitch and loudness can of course be considered as targets for machine listening models, but more importantly, interestingly and usefully, other qualitative features can be discerned to distinguish sounds of a particular noise quality, say, or with particular areas of spectral energy density, and to track some patterning of these sounds across longer segments of time. To this end, I have spent much of my time in the Institute of Sonology developing listening models with digital signal processing methods based on the instantaneous (moment by moment) frequency of the input signal, the autocorrelation function of that input signal, and the spectral distribution of its energy concentration on different time-scales. These will be discussed individually in Chapter nine. Further, a great wealth of information can be gleaned from a working method that combines several different analysis techniques, allowing control signals and sequences to be crafted to suit each desired situation of interaction.

In every type of analysis, we must, at least on some level, deal with the concept of size. Douglas Kahn relates that **"the practice of considering sounds according to size ... appeared [in music] at the turn of the century, when it became apparent that existing means of musical notation were inadequate to the tasks of denoting smaller and smaller intervals and of representing many of the salient characteristics of sound in general"** [Kahn 1999 p193]. While Kahn is here referring to concepts of pitch intervals (after all, classical notation has no significant problem referring to smaller and smaller intervals in time, simply adding another 'line' to form hemi-demi-semi-quavers (sixty-fourth notes) and by altering the overriding tempo marking when this method becomes illegible), most machine analysis techniques rely instead on sizes of the amplitude or energy variation in the signal registration. Loudness, related to amplitude, has frequently been ignored even in electronic composition. Considered somehow of lesser importance to an otherwise precisely organized construct of pitch and time

information, any fixed media work be it on record, tape, CD, etc. still relies on an uncomposed and uncomposeable amplification process to bring the results from the vinyl, electromagnetic or digital domain into the human hearing range. Therefore, it can happen that although the relative loudness of the various constituent elements of music have been agonized over at one particular amplification level, that listening more quietly or loudly destroys this balance and leaves the composer wishing to remix the work again.

This area of discussion highlights a particular realm of confusion, namely that **"when a particular representational framework becomes the prevailing means for producing and observing a signal, cultural practice often collapses the identity of the signal with that representational framework ... However, through acculturation and education, we fail to notice this and instead treat such a signal as though it were in fact "raw data" ... Through habit and acculturation, we become unable to observe the descriptive and representational apparatus by which a signal obtains its coherence: signal and representation are subsumed as one"** [Hamman 1999 p91].

As a machine is only capable of listening to audio signals from a microphone through an analogue-to-digital convertor, any machine listening model must begin its life in this way as a digital representation of such data - energy distributions through time. The technique of describing sound as such a vibrational pattern has revolutionized the way we think of music. Quite independently from the musical problem of displaying tiny pitch intervals on a hand written score, technologies infested the modern world and advanced the instrumentation available for our use in exploration of sound. The implications of such technologies are still increasing their impact in our musics, as for example with **"phonography [which promised] an alternative to musical notation as a means to store sonic time and, in the process, deliver all sound into artistic materiality ... Musical discourse responded by trivializing the complexity of significant sounds and their settings. Indeed, after a certain historical point, it was not so much the potential for musical practices of imitation that were debased as it was the concept of imitation within musical discourse"** [Kahn 1999 p103]. Concepts of imitation include elements of repetition and therefore relate to one's recognition and memory and hint at something truly fundamental in one's experience of sound and music.

In a significant step further towards the eventual use of data from machine listening models in an interactive environment, it is helpful to consider some audio analysis techniques that, while they may not provide anything to

listen to, do provide some potentially useful information at output. Often the output is to a visual display, re-definable with parameter values moving on particular dimensional axes or scales. Many other signal representations can be derived more-or-less mathematically from the basic audio signal that describes air pressure fluctuation over time and can be measured by a microphone. Perhaps in the future, research projects might rely directly on a brain analysis allowing one to dip straight inside one's own mind and project or produce one's inner sound imagination straight-away for another person, whether via the physically sensed world, that is, the loudspeakers in the air vibrational 'delay chamber' of the world and our perceptual processes of reproduced audio, or whether somehow as a hallucinatory experience straight to the listener's mind?

For the meantime, however, we stick to audio alternatives that are available. The experience of audio analysis most likely to have been met by the general public is something akin to the game "Karaoke Revolution" developed in 2003 by Konami for a popular home games console. It is a multi-player game that uses a microphone and voice analyzer to listen to the gamer's pitch and rhythm while they sing along to popular tunes, delivering a score-mark depending on the accuracy of imitation. A visual feedback system displays a player's progress in a timeline across the game screen.

A system such as this would be easily achieved by taking a spectral analysis method that measures energy in particular frequency areas (eg. the Fast Fourier Transform) in conjunction with an amplitude envelope ('following') function that has appropriate threshold levels set so that when a player is singing loudly enough and within the correct pitch range, their cursor or avatar in the computational model can be manipulated. This particular game plays by the rules of octave equivalence, being insensitive to the differences in male and female voice registers and thereby allowing both sexes to participate equally. The analysis underpinning the game provides an example of 'score following', **"a pattern matching technique that traces the progress of a live performer through the composition [he or] she is playing"** [Rowe 2001 p212].

In the composition of an interactive installation, an environment that has often been seen to be cross-modal in nature, it is useful to have a number

of separate controls 'on the tip of one's tongue'. However, these dimensions are cross and are overlapped, provided in an intertwined web open to various interpretations. The above game discussed the singing of different pitches, but what 'pitch' could be detected without a 'positive' reading amplitude or presence of energy? We cannot measure a vibrational frequency pattern without there being some movement itself that becomes measurable also. That is, there must be sound present in order for it to be said to have a 'pitch'.

In this way it is seen that certain dimensions of audio commonly held as 'separate' are in fact intrinsically linked. Traditionally, a notated score uses a vertical axis to represent pitch and a horizontal axis to represent time or duration. Time always moves perceptually forwards, whether this manifests its represented as left to right as in a score, or from top-right to bottom-left as is the case when working with the compositional conventions associated with "MaxMsp", a programming environment developed by Cycling74. Whatever representational boundaries are selected as 'grid', time is re-stretched within the boundaries that the conventions of musicality have extended, and a set of symbols are placed on it that are deemed to be representative of compositional decisions.

Audio analysis methods have, like synthesis methods, been available for some time. Early techniques were carried out by acoustic researchers, particularly in the quest to define the term 'timbre'. Having separated amplitude and frequency in time, and the term 'spectral centroid' was not long in coming as an aid in describing the element of difference in sound quality between two tones of equal pitch and loudness. Processes of statistics and modeling could be used to elaborate real and artificial signal chains and networks, affording inspiration to composers such as Varese, for whom the timbre of sound was a more basic structural component of music than melody, harmony or rhythm.

Through longer periods of time, **"perception is a process that has evolved to interpret the statistics that exist in our environment. Things that are common and repeating we tend to subliminally ignore, and things that do not make statistical sense confuse us (auditory illusions, noise)"** [Smaragdakis 2001 p102]. This is the realm that installation artists can choose to work in, creating in time their own science-fictions,

manipulations and experiences. Thus a particular musical environment can be engineered allowing the human and machine parts to blend into one sonic world.

More compositional and technical information follows in Chapter nine with regard to my recent activities both practical and theoretical (to give them their full 'title') the digital-audio-signal-processing-analysis-models that I have programmed during recent studies. These 'listeners' are like little birds waiting to be fed - each with a character of their own. Each waits for particular kinds of input (audio, or at least derived thereof in part), and judges the input signal according to specified sensitivities in behaviour and will react as its state or 'character' decides. The remainder of this Chapter outlines the working methods used to accomplish the 'machine listening' aspects and audio detection as required in this work. Works of others are discussed below.

Making a broad generalization, the task is one of "interactive music summarization". Research based on such a generative theory of tonal music has been undertaken in Japan [Hirata and Matsuda 2002], seeking to (automatically) find the most distinctive representative musical excerpts. They follow a two-sided approach to analysis, one approach from point of view of the audio signal (considering key phrases, repetitiousness, memorability or not, etc.) and one approach from point of view of the symbolic description (the level corresponding to a score).

Similar research has been undertaken to define a 'music snippet' in regard to the most representative highlight excerpt of a music clip, similar to image and video thumbnail [Lu and Zhang 2003]. The frequency and energy features of the whole music is considered in order to determine salient segments, and each of these in turn constitutes a time-segmented analysis sensitive to boundaries around individual phrases. A selection of possible sensitivities allows this parsing of the input data to occur, and furthermore can be considered in order to attempt to avoid (or to match) a clear beginning or end point and instead start mid phrase, mid sentence.

A system of analysis somewhat similar to the "Karaoke Revolution" software previously discussed (but now with a much more sophisticated way to deal

with longer units of elapsing time) is described as 'query by humming' (QBH) [Song et al 2002]. Their work involves a melody feature extraction process to track the dominant pitches, based on the fact that melodies can be formulated as the consecutive sounds that are most audible and predominantly pitched. Looking for pitch, they seek clear harmonic structure in a sound with large amplitude. Other considerations include masking, segmentation, strength, silence by segments - all techniques that rely on quantizing and re-reading portions of time and frequency analyses of something sung into a microphone.

Based on the QBE example, the user in the 'sound spotter' [Spevak and Favreau 2002] selects a particular reference passage and the system then retrieves 'perceptually similar' occurrences, defined as similarity of spectral evolution (measured by comparing sequences of feature vectors). Displaying a pointedly current fact, this work among many was designed primarily to evaluate its own set of methods, here retrieval algorithms. Four stages of work allow for frame based feature extraction, clustering, pattern matching and a consideration of the number of best matches to relate. Similar pattern matching techniques are described elsewhere that allow greater flexibility for data structure descriptions by allowing also for mis-matches and dynamic changes to occur within larger structures [Malik, H. 2004]. This paper is situated in a purely computational environment but nevertheless examines concepts useful in the composition of sound and sound control in interactive audio installations. The concept of 'edit distance' is used to compare the efforts needed to transform a source sound into a specified target string or sequence. Many such features are implemented in the MaxMsp patch "ssshahee--" presented in the Sonology Colloquium series [2004] as an example of a vocal-timbre recognition system that reacts to occurrences of ss, sh, sh, ee and the general-background-sound present with various modes of sensitivity showing in its output displays.

Precisely because we as people experience changes of things in time, we desire of our machines that they become sensitive to these moments of change in our perception. Within a computational software model, a person can in part be considered in mathematical terms in part as integrator and

differentiator, influenced by certain principles of listening, and acting according to particular cultural statistics and psychological pre-dispositions. When questioning what our musical intentions or even interactions might be or become, we would do well to remember some perceptual groupings that occur. Principles of commonality are sought, such as common frequency, amplitude modulation, onset time, harmonic relationship and spectral shaping etc [Smaragdis 2001].

We find also that these patterns might be perceived as some commonly modulated statistical property and therefore that streaming clues such as closure, continuity, and pattern repetition might further affect the component group's chance in exhibiting common fate, or not, and therefore of being perceived as one sound object, or not. A great deal of further research has been carried out into, for instance, sound source separation and scene analysis, another technique that relies on grouping characteristics. Other techniques aim more to model sound as a process, constraining time-scales of investigation in order to expose specific behaviours. Often limited at present to working in small chunks of sound (often only a few hundred samples per 'window' or 'buffer'), analyses must often be drastically re-made to delve from one speed of action to another. Very many more techniques of sound analysis, stretching across frequency, time and other domains of consideration are implemented as self-contained algorithms aimed at for our perceptual relation [eg. Jehan 2005].

Implicit in all of this work, but yet unstated is the view that such analysis techniques yield signals of some substance, that, despite the lack of representation of modern theories (musical or otherwise) on auditory perception, this information is of a form that can be used to further musical purpose - **"through the intermediary of sound, digital data is figured here as exactly the thing that it is not: matter"** [Whitelaw 2003 p93].

chapter seven

Kevin Warwick, a man who has been living since 1998 with technological implants (sending radio waves between his human and computer centres), has realized that **"when something is fixed inside your body as an individual you regard that thing as being just as much part of you as your own bones or brain"** [Warwick 2001 p43]. A major benefit he sees lies in the incredibly far-reaching power of the fact that **"[whereas] the human brain has evolved to think in three dimensions at most, sometimes only two or one, machine brains can think in an almost infinite number of dimensions"** [Warwick 2001 p42]. The ability that we have as humans is to spot the patterns that make our impressions make sense. This happens automatically and without analysis, but tempered with our goal as humans to improve the communication links between each other and with machines, we find it difficult to define how a machine ought to do this. We have, in the previous two chapters, seen some ways in which a computer environment can interpret its human and contextual inputs whether in audio or other signal domains of representation.

So, we want to direct the work so that particular analysis tasks can be carried out in variable configurations suiting the (musical or otherwise) moments at hand. The input and re-processed audio and control signals can, for instance in the MaxMsp 'language', be sonically analyzed through time so that derived audio and control messages may be sent, by means of their having been correctly defined in influence over the appropriate decisions, at the appropriate moment, to whichever listening receivers are desired. Working furthermore in a recursive programming manner allows one not only to model the individual relationships between any signals as ratios or proportions of each other, but also to recombine these units again and again, re-set over and over, placed one within another and so forth.

The world today seems to support and encourage the stance that participation be kept open for everyone, and that musical participation is to be promoted and rewarded. In such commercially competitive times, service-providers should maximize their chances of success by aiming not to lose one single customer. The artist through his or her tools can reflect this benefit by keeping an interactive work accessible to as many people as possible, and yet develop a sense of musicality, as it were, in the widest sense possible, by encouraging an exploration in sound that can be varied and influenced by actual sound created live, within the character of the

work in its specific installed setting.

As many questions in this field of art surround the idea of communication - of structuring ideas, words, methods, messages, sounds - the human voice has now become of fundamental issue to such works. Another factor underlying these work reflects on the development of a critical ear that is necessary for all musicians, and the process that a musician undertakes of practice, creating repetition and variation. Bearing in mind the principles of our perception in terms of grouping and separation, recent models have awareness of some statistical prediction methods concerning particularly defined movements across scales of interaction between the interconnected audio-analysis units in the environment of their actual installation.

Across all aspects of life, **"speech technology is getting increasingly better, cheaper, and easier to work with. Interface researchers predict that speech will be the major form of interface in the future, especially as intelligent devices diffuse into the everyday world. What will these devices sound like? What will be the limits of their ability to understand us?"** [Wilson 2002 p784]. Although frequently working with the human voice, many models used in musical investigations hold to the fact that **"a word [is] too different from its sound to be unambiguously transcribed"** [Kahn 1999 p92] so, while being influenced by semantic concepts compositionally, restrict the components of sound to be recorded and interpreted according to sonic rather than semantic detection methodology. The following chapter of this thesis documents some digital signal processing (dsp) models created recently that deal with live audio inputs from microphones, viewing interesting features of sound in the coordination of, for example, pitch-to-noise switches and template matchings that allow vowel and consonant patterning sequences to be observed.

Two types of voice recognition commonly discussed in the world are speaker-independent and speaker-dependent technologies that aim to recognize the linguistic content of a spoken phrase. The former describes a system that must be trained to recognize a particular voice and currently can achieve a vocabulary of around a quarter of a million words. The latter method has a far smaller vocabulary of known words, but benefits from the fact that it need not be trained to individual users and is thus approachable by any speaker. However, this chapter introduces another idea, primarily that of

using the voice in a non-instructive manner, focusing directly on sonic rather than semantic content of the utterances. Hidden beneath the veneer of such techniques are hidden implications that surround their usage, specifically that one speaker be present at a time (for a kind of monophonic transcription process). Various methods of feature extraction are applied to the voice in these techniques, and increasingly, are being exploited in commercial use. We find voice command recognition implemented within the operating system of many computers now, voice print recognition used by the police and intelligence services, and we are now also beginning to observe results in the artistic and musical creations the world over.

Avoiding purely referential techniques and iconic imitations of particular events, we can instead treat the voice exactly as we would treat any other sound source before our ears, looking for the interesting sonic features within it. Rhythms of spoken text limit our perception of whether or not we hear any linguistic sense in the sounds emitted, and the voice can therefore be used in non-linguistic but musical ways, still perhaps singing or talking but dealing now with concepts more basic to communication, passing messages, exercising play-forms, restructuring phonemes, alongside ideas of mistakes and impatience and so on. From this perspective of analysis, the content of semantic message will frequently be lost, but its style of delivery may be over-emphasized to compensate.

A person may be encouraged to find that they are able to use their voice in ways other than the usual spoken communicative methods of everyday speech. Reported on the BBC news website in 2001, it was stated that grunts and sighs, ahs and umms could become the way to control computers [Igarashi 2001]. Rather than extending techniques of moment-by-moment recognition into longer units of time by seeking word-units, or letter-lists at least in the way that the individually recognized units are sequentially combined into longer phrases, Takeo Igarashi's work in this area relies on the insight that a system based on simple sounds instead of words would be more efficient than conventional voice recognition software.

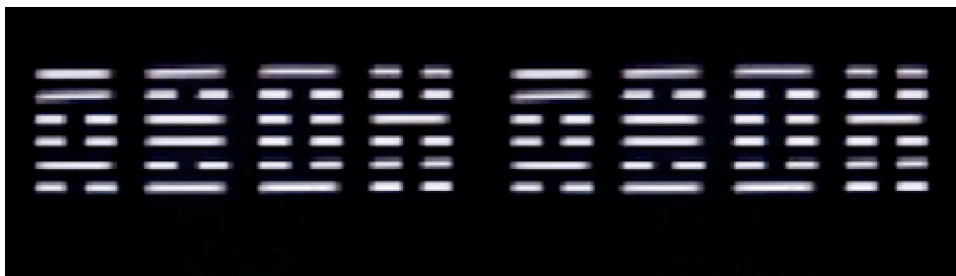
This is also the theory that informs collaborative work towards a sensitive stage conceptualized by choreographer Maria Ines Villasmil. The first incarnation "Ecce Space" (May 2004) posed the problem of creating a voice-

controller to react to a long spoken "sh" sound and fade the lighting of the space to black. A second incarnation of this project, "RECOG" (July 2005) worked on during a short residency at STEIM, resulted in a work involving a system based in motion tracking, voice command and (target proximity matching) timbral-recognition processes to allow the dance and the environment, together, to present the seated audience with an integrated picture in sound and image. Semantic issues of vocabulary and identity (perhaps, say, the recognition of a voice within a crowd rather than of a particular feature within a single voice) are not directly addressed in the programming requirements of the system, yet the issues are implied to the attentive or reflective members of the audience in the presentational context of the work - a pre-composed staged performance where many elements behave 'live' under the joined control of person and machine. The active listener is able, through considering what they are actually hearing and otherwise experiencing, to correlate particular vocal actions with associations in other media, and to listen to how these associated relationships then alter during the course of the piece.

Chapters six and nine will describe some models developed to deal with sound-inputs from a microphone, a development of the representational display techniques outlined in Chapter five that can often trace their conception to ideas in scientific research. Despite the audio domain of control, many different media can be controlled with the resulting signal models, and indeed the model itself can take many forms of structuring. The work described just previously into a detection of "sh", for instance, will be seen to have arisen from a model of the autocorrelation function of an input signal (in Chapter nine). However, to detect evidence of a sibilant sound this could equally have been achieved using the root-mean-square amplitude measurements in combination with a method to count zero-crossings in the signal representation. This is in fact the method selected by Zack Settel in order to allow **"the vocal quality of the singer's performance ... to affect the synthesis and processing of the electronic part"** [Rowe 2001 p224-225]. This area of work provides an area to explore the 'sense' or otherwise in musically offered words, playing on aspects of direct-copy and distortion as vocal aspects come in and out of our perceptual attention.

When faced with a situation of some activity or sensory stimulation around

us, as humans we set our minds to the situation, wondering what is happening. The picture which follows is a still screen-shot taken from "64s", a work made in collaboration with Georgio Bachtsevanidis and presented in various stages of development in Rotterdam this past year. An audio visual work, the patterns on the screen are altered according to binary sequences (6 lines vertically stacked represent one of 64 possible individual states). An interesting observation was made that on seeing these patterns most viewers would report (within a few seconds, even) having seen letters, having read words across the screen.



A factor analogous to this is seen, or rather, heard, in electronic music when a vocal element often becomes the thing suddenly most easily recognizable in an otherwise incomprehensible texture of sound. Repetition is also a factor crucial in this recognition, and we are reminded of its importance as an element in compositional choice - a fact largely due to the influence of the very technologies in use. Douglas Kahn sees in the period since the 1920s onwards, that **"against the obstacles of musical thinking, the case for auditive imitation became increasingly compelling, and with it came a new sense of artistic possibility, a marked increase in theoretical and practical activity in the artistic use of significant sound and important experiments in asynchronous film sound. Unfortunately, radio art, audio art, and film sound experimentation based on recording technologies was cut short and postponed for decades"** [Kahn 199 p123].

The ability to record amplitude variations of sound pressure waves allowed many such aural abilities of humans to be explored, calling into question how our perception of structure can be influenced. Today we have many other

techniques of registration available with which to attempt a determination of audible structures in the sounds and music we hear. The remainder of this chapter outlines findings of research into other pieces of information that can be gleaned from such processes of signal manipulation and representation.

Words, like imitative sounds, give rise to associations in the listener's mind and as such have been problematically viewed in areas of music and science. Setting the scene for such investigation and highlighting the difference in sonic and semantic ways of working, researchers in the twentieth century focused on Multi Dimensional Scaling (MDS) methods of representation to model sound. Originally such works were beset by criticism over the fact that the experimental process was flawed, its design showing the bias that the results had - a reliance on linguistic association. Participants in one early study were asked to judge sounds artificially removed from any musical context by verbal categories of distinction such as bright/dull (investigated in 1941) and later with regard to their qualities of sharpness or roughness (investigated in 1974) [von Bismarck 1974]. Clustering of results on axes or particular regions were sought as perceptual correlates to **"the orderings of stimuli along the various axes or dimensions"** [Grey and Gordon 1978].

These methods were then somewhat of such associations of linguistic terms by introducing more simple perceptual scales - those of similarity and difference [Miller and Carterette 1975]. Alongside this change, more flexible temporal scaling aspects were introduced to the model, allowing a greater consideration of transient aspects of sounds more true-to-life than the artificial steady state sounds used in early experiments. However the fact still remains that when a multi dimensional scaling approach is taken to analyze the human way of listening, one will only ever find results in terms of those dimensions with which one actually looks for them. These comments highlight the context-dependent nature of such analytical works, and show that care must indeed be taken of one's own predisposition towards particular sound description and data reduction techniques.

Music is often said to be an international language of communication simply because many things can be achieved in music without recourse to literal

definitions, as has been suggested in the discussion of the collaborative working methods often adopted at present. Aside from referential and other implications arising from repetition and so on in the sound or music under question, many audio analysis techniques are in use that rely on some aspect of our nature as human beings to make meaning from recognition of similarity and difference as experienced through time - namely change. Scheirer [2000] has suggested a method based on spectral shape hearing where the differences between spectral input representations are measured by subtraction, and his work addresses subjects such as the return of previously heard material, the arrival of new material, and changes of texture.

Relying on psychological methods of experimentation, an idea of what constitutes a typical human response to a situation is implicitly embedded. Joseph N. Straus points out to us that this has arisen in part due to our abilities to statistically model a situation, and that **"prior to roughly 1840, the relevant constellation of words (normal, normalcy, normality, norm, abnormal) had not existed in any European language or culture"** [Straus 2004 p22]. Issues of repeatability and repetition rates are intricately tied up in our abilities as humans not only to make independent snap decisions about recognition, but also in regard to being able to match one's current situation with that of others, and includes one's memory of one's own prior experiences.

We can now control many more dimensional elaborations along and between the electromagnetic, computational and perceptual scales than ever before. One of many possibilities arising from such a scaling, of the perceived mallet hardness for example, was discussed earlier [Freed 1990] in Chapter six. Rather than base current methods in the western tonal perceptual tuning system (which cannot describe the developments in timbral and structural control of sound afforded to us by the instruments of recent musical and technological developments), other scales of relational importance to our experiences as people can now be examined.

chapter eight

The crux of the situation concerns **"the cultural imagination of data [which] is crucial in a society increasingly enmeshed in the datasphere"** [Whitelaw 2003 p93]. This chapter of the thesis re-introduces the overlapping areas of technological and musical innovation (linking back to points raised throughout this thesis), primarily showing how areas of scientific research are rising in importance to sound-practices and are playing an emphatic role in the current contexts of music in the traditions of the western art world.

Stephen Wilson, in his report on the interactive arts relates to us that **"much early artistic activity with computers focused on their capabilities as manipulators of images and sound, probably because these were closest to historical forms of the arts"** [Wilson 2002 p605]. He thus outlines a person-biased, tool- or instrument-using conception of interactivity, as is held for example by software designers, games and immersive virtual reality creators, a view that that concerns itself with solitary interfaces and user interaction communication aids. He continues by noting (albeit from our visually biased stance) that this conceptualization places the use of computers in art **"in a direct line stretching from Renaissance perspective through painting, photography, cinema, and television, to the present moment. In this analysis, the computer is not a radical breakthrough but rather a continuation of narratives emphasizing the unitary point of view, the controlled frame, and the manipulation of desire through image"** [Wilson 2002 p605].

We must note, Wilson continues, that **"it is a mistake, however, to focus on only one area of activity in trying to understand the underlying meaning of computers, their impact, and their relationship to culture and art. Computers and their associated technologies represent the culmination of many techno-cultural streams."** [Wilson 2002 p605-606]. Whilst the implications of computers to musical possibilities in the present are of course as yet uncharted, unimagined as they are, we can presume that the effects of their introduction will be enormous on how a human may experience music in times to come, exactly as technology's impact has been felt in all aspects of daily life.

This thesis directs itself towards those methods that maintain the 'social' elements in such works, bringing people together to share their experiences. As such, it is not concerned elsewhere with the distance-based possibilities that arise frequently, seen for instance in internet-based

'net-jamming' improvisation possibilities. Gleaning insights into the 'dynamic' of a live situation has led to developments in musical instrument-making and playing processes. Human nature predicts that it is almost inevitable that the action-at-a-distance home-based individual approach, where one plays by IP and the internet protocol system thereby governs what information can be relayed, will lead too to new developments in musical composition and research. Indeed, this method of working is frequently used by composers in the way that files can be shared between collaborators and colleagues with the difference that in this case the moment is not of the real time performance distribution 'moment' considered 'the music'.

After further development in work on body/mind-machine interfaces has allowed us to progress further in technological doings and understandings of necessary data conditioning and pattern-spotting processes as they relate to a human's perception and intention, we may be in a better position to consider another implication which will shortly hit home. Without requiring sensors attached to an instrument or such context-based paradigms of performance data, we may become able to 'somehow' read another musician's body language (more pragmatically their 'controller data') but from a distance, whether visibly, aurally or representationally construed. Data-reduced signals, whether from such audio, video or other sources, will be available for direct broadcast not only globally, but also locally, to other musicians or participants present within the same system. The implications of this, too, are being found in many areas of life, and again an example can be made in military research where work is being undertaken on the feasibility of remote telesurgery to aid soldiers in the field.

When a player performs (on) an instrument of any kind, control data can be said to be produced. An acoustic instrument will produce audio data primarily, but one could likewise consider MIDI or indeed any convention. When we allow that this information can be broadcast, we encounter further questions in music - ideas of personal space, audition ranges and so forth. We must consider, for instance, what may happen if a person's data trail can influence several machines, and whether they are acting consciously in a particular manner or whether indeed their data movements are being siphoned off unbeknownst to them, stored in a surveillance module and

passed from there to others [an example of this in the visual domain, "Watched and Measured", based on security camera footage of visitors entering the building, was installed by David Rokeby in the British Science Museum, London in 2000].

Many such topics interlink, now, as processes that were once only compositional become 'playable' in the live moment thanks to the advances in our technology. Similarly, through analysis techniques and a wide variety of monitoring methods that exist (and will continue to be developed), we find that details of sound that were once lost in the passing moments of a live performance can now be examined and noted, composed.

When dealing with music, we have as humans evolved a set of conceptions based on the tools that were previously at our disposal, naming for instance the use of instruments-for-making-sound and instruments-for-storing-text as separate acts of playing and writing (or typing). Thus our development as a technologized species has already influenced the music we make up to this point in time. Specific examples of this can be seen throughout music's history, as when for example an extended instrumental technique becomes apparent, exploited in the composer's output of that era. In recent times we have all developed high levels of skill at 'pointing and clicking', but what use is this to music? How would we rather 'transmit' our musicality?

With these questions, our search becomes one of **"the translation of the world into a problem of coding, a search for a common language in which all resistance to instrumental control disappears and all heterogeneity can be submitted to disassembly, reassembly, investment and exchange"** [Harraway 2002 p302]. Robert Rowe warns us, however, that a general representation for music does not exist, stating rather that **"in order to represent something, its properties must be interpreted according to some proposed utility ... One cannot meaningfully discuss the design of representation schemes without some knowledge of how such a representation is going to be used"** [Rowe 2001 p31] and that **"a physically based sensing mechanism should be able to capture information corresponding to the available degrees of freedom"** [Rowe 2001 p215].

Ideas of uncertainty in relationship abound in the electronic music world, echoing perhaps the technological and conceptual developments implied in

(and beyond) relativistic thinking in science. It seems that those who control the technology control the aesthetics surrounding it, yet it seems impossible to have an aesthetic understanding of an interface alongside a working version of the object itself: everything learned from one instantiation is poured back into the next refinement of itself. Each experience builds on top the previous, and understanding gradually dawns.

In an interactive audio installation, each player plays their own part within the environment, there, presently. Each person contributes what they will, and the work continues. The point of artistic work about interaction then becomes one of interpreting which situations are 'valid'. However, as Chapter seven makes clear with respect to MDS, as Chapter four relates with collaboration, and as Chapters five and six show with respect to movement data and audio signals respectively, we are, yet again, confronted by questions of literal (verbal or mathematical) definitions and cognitive meanings. The difference now, however, is that whereas other problems were seen perhaps as technical problems to overcome, we can now work on an 'artistic' compositional level of decision making. Thus we question - for how many people is the work intended? Is it aimed for musicians or non-musicians? Will it allow interaction? If so, by what means? Does it allow some exploration of temporal dimensions, say by controlling and remapping live input sounds according to this input within its current (and prior) conditions? Does the work allow a transfer of energies from some areas to others?

When we consider installation art works of the interactive variety, we find that no matter which areas of sense-representation are appealed to (and derived thereof, for example in sound, silence, loudness, pitched, rhythmic, timbral or other structures), examples of mutual dependencies are present. In such works, we are not striving simply to set up a condition of machine listening, but rather to supply sounds with energy and if desired to give them a life within a particular context. When working with sound, therefore, the listening models must still be 'mapped somehow' also to the final audio units that produce and direct the eventual sound out into the environment. Thus we find, **"adjustments in the interference among sonically relevant parameters ... may be the object of design, and hence worked out creatively as a substantial part of the compositional process"** [Di Scipio 2003 p271].

On a compositional level relating to both sound materials and structures, I am particularly interested in sound works whose musical content and structure are conceived and produced with reference to concepts from technology, such as with repetition and sequences. Rather than the idea of music as existing in one pre-ordained sequence, I instead envisage a state where the music 'has' or 'is in' the potential to go in many different aural directions over a period of time. I, like many others, use various synthesis techniques to explore concepts of time in these works, attempting to concentrate these in such a way that the player may effect changes within a framework that is, to some extent, pre-composed in both the selection of certain possible preset states, and in their associated interrelationships.

Aiming to increase people's engagement with the work, some people have criticized the fact that interactive work often use gimmicks to prompt action (eg. the use of the 'doorbell' metaphor in "Summoned Voices", an installation discussed in Chapter three). Such tricks do have an important function, however, that should not be automatically scoffed at or overlooked - a bodily involvement with a work will improve the chances of some mental involvement occurring also. Many people exist in somewhat of a daze, so stressed out by daily life that they cannot 'let go' when they have the chance and instead may spend time (even during a staged performance) worrying rather than listening. Though they may have arrived without the inner-composure that allows one to bring one's mind into a room when one enters it physically, the artist can if so desired then aim to re-attach a person mentally with their actual current experience, can perhaps persuade them to slow, relax and undergo a new experience of sound. When discussing interactive situations then, no matter what actual data is being observed nor what media are used for output, we find that artists are aiming to create potentialities of certain moments of interest being experienced, and are sharing some power with the interactor to bring about varying transitions or articulations between such moments.

Perhaps working along the lines of compositional algorithms typical to interactive systems (Rowe mentions three - sequencing, generation and transformation) we find that we must rely on **"control strategies ... [that] concentrate on**

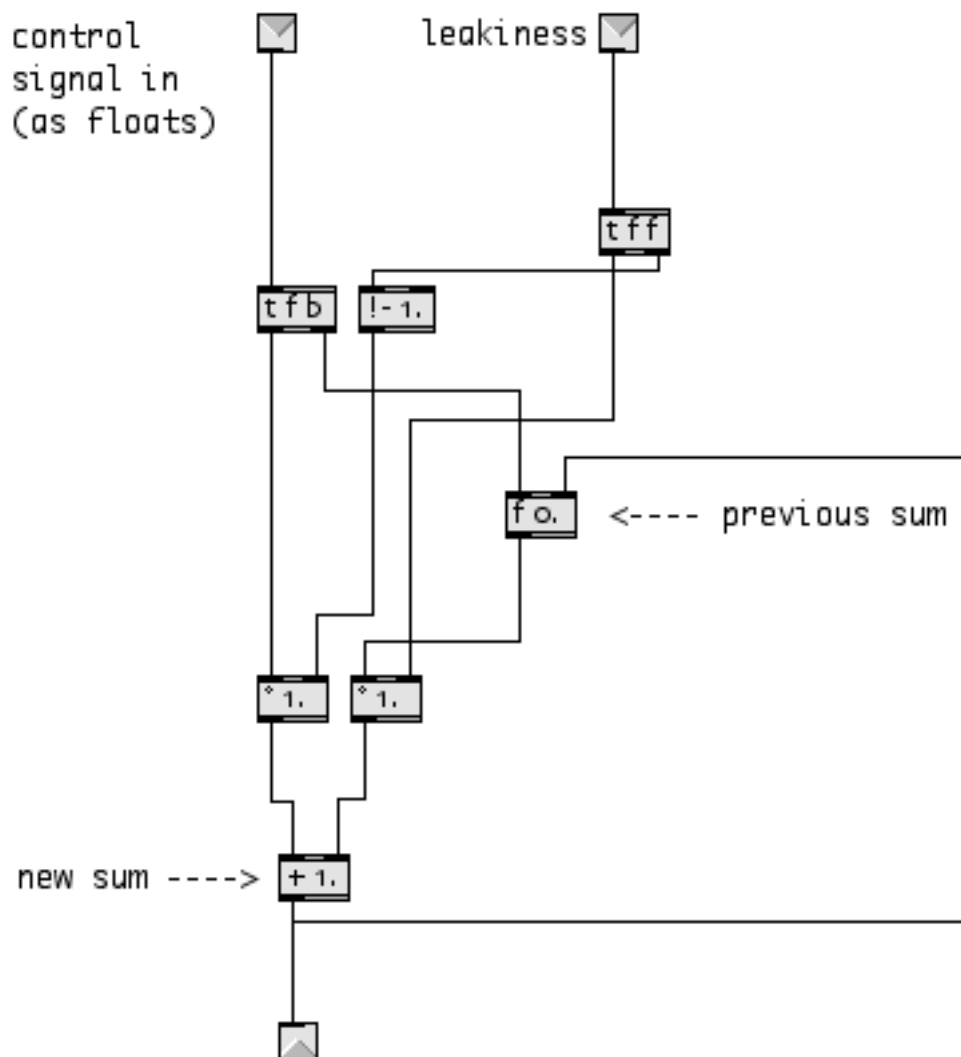
boundary conditions and interfaces, on rates of flows across boundaries - and not on the integrity of natural objects" [Harraway 2000 p302]. This idea is reflected invisibly in almost every technological item, and is implicit in the very nature of human activity. Any process of following, averaging, summarizing, etc. will also be concerning itself with such rates of change. We see this point in its fundamental form in the "leaky integrator" discussed below - this is perhaps the most basic idea absorbed from scientific practice into musical habits and allows us to deal, in varied manners, with the concept of change.

The leaky integrator concept describes a device that 'smoothes' the input signal in one of several ways. This method is one of the starting blocks in many pieces made and as such is described here in some detail. An implementation of this idea in the MaxMsp environment is also given on the following page for those who are interested - this version relates to a control signal, a stream of numbers arriving at the top-left of the diagram, but the technique transfers through to other domains of representation.

A way to smooth the signal received at input uses an internal 'memory cell' or feedback system as follows: a value of zero leakiness allows none of the output produced to be reintroduced at the input alongside the current live input value, whereas a value close to one will allow nearly all of the currently produced live value to be 'fedback' into the system, reintroduced (as an altered version of the previous input) at the currently 'live' inlet. In signal domain mathematics, two signals 'on the same wire' (ie both attached to the same signal inlet in a max-msp patch) will be 'added together'. The value of the feedback coefficient thus influences the sensitivity of the unit to dimensions of elapsing time. A zero leak coefficient implies that the output will always respond instantly to current live input (having no memory even of the just bygone events). On the other hand, as values for the feedback coefficient increase towards one (approaching from below) the 'memory time' increases which slightly delays the responsiveness of the output to change in input and thereby smoothes out the input curve. When the coefficient is exactly one, the value at output remains stationary and a 'snapshot' can be made.

integrator circuit, in control rate domain

every time it gets a number, it gives a number



We can make significant advances in a great number of audio analysis and synthesis methods with this trick. At any place where a control signal could fit into a larger scheme, it is often useful to have a version of that input that can be smoothed out before being passed subsequently onwards and into the next layer of the software. Further, damping methods may also be selected variably to allow us different manners of translating the signal's response, for instance whether it reacts evenly with equal distance moved in each direction of distance from centre or whether it responds more quickly to movements in one direction than in another. Thus we could in a very simple way make, say, a flow of pitch information that could roughly follow an input pitch but that would either slowly swoop both up and down, or swoop up but step immediately down on any lower input, or indeed step up and swoop down.

With the power of today's computers, it almost goes without saying that we can use such a copied version of a signal in place of the original and to the same purpose. For example, compression effects are often used on a vocalist's microphone to reduce the overall dynamic level variation into a smaller range, limiting the 'change' allowed in the signal. In terms of music with electricity we continually use representational copies rather than originals of signals - we dealing with templates, functions, algorithms, procedures and methods, applying one thing to another, modulating one with some function of another and so on.

A philosophical extension of this issue has already been met (in Chapter four) in the commercial audio world regarding intellectual copyright and theft, especially with respect to issues of sampling in live and non-live electronics. However, it remains to be seen how such issues will be resolved in a world increasingly removing the physical media that have traditionally been associated with music - the score, tape, vinyl, cd, digitized as .aiff or .mp3, etc. all fight to claim music as their content, yet streaming-audio-players already retain no copy of the sound material once it has been played and there is, as such, no carrier of music to speak of. The potential of the listening experience is the consideration that rules such questions at present, whether in live, installed or transported on a fixed media format.

chapter nine

Rather than rely on commercially available packages for sound analysis, I have, during my studies in Sonology and in a quest to broaden my understanding of sound, developed various listening models in the MaxMsp environment. These techniques are implemented within separate modules and can, true to the methodological considerations of interactive art practices, be recombined in many ways. Each reconfiguration allows access to different portions of a sound, and those portions of 'significance' to a particular context can therefore be sought.

By analogy with methods of communication central to interactivity, each of the digital signal processing units contributes to a language in which to listen and speak. Each unit, perhaps like a word, can thus be combined with others in a larger structure to create phrases or sentences - data structures with a particular characteristic or intention. Further, altering such combinations and modulating their internal connections through time (with a combined effort made by human and machine conditions being satisfied) we thus allow longer structures, stories to be told.

Whether we chose to conceptualize a behavioural characteristic to the computer's function in the environment (eg. that the machine fuses with, opposes or ignores the human) or whether we dream simply in audio terms themselves, we remember that interaction has been best described as systems of interdependencies within the individual components. Four such components, listening models named acf~, if~, fsum~ and reel8 will soon be introduced.

These models accept digital audio representations of the environmental sound at their input, and derive both audio signals and control values that are passed from their output, either arbitrarily- or perceptually- matched to another process that may produce or control audio directly or any other media (lights, image projection and so on). Working individually to report judgments on the quality of the audio signal under investigation, these models can, in combination with each other and with an appropriate control structure, thus allow the input sound to be used in various conceptually different ways across time. Decisions on the control structures to use form the major compositional tasks.

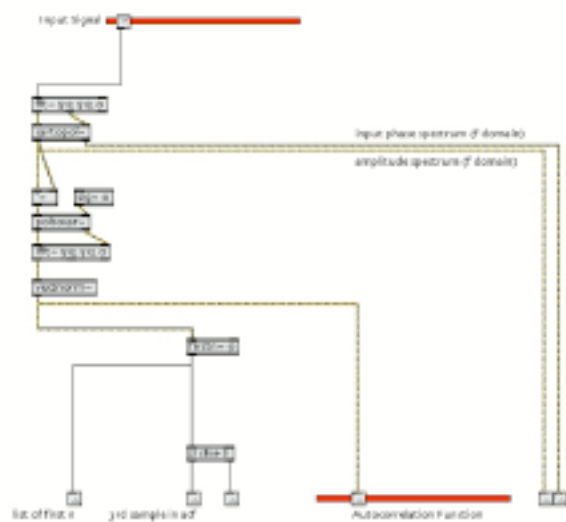
Control structures for interactive installations can be based on purely musical decisions and intentions, or can be physically modeled, for example, on patterns observable in nature or the technologies of mankind's devising. Indeed, they may be structured on 'if-then' algorithmic sequences as described by Joel Chadabe (and discussed in Chapter three). They may be as 'arbitrary' as one can imagine.

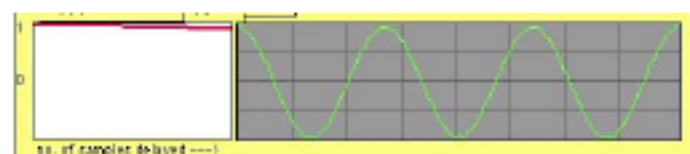
Strangely perhaps, the ancient Chinese philosophy contained in the 'Book of Changes' has proved useful in devising such networks of interconnectivity, allowing a definition of how change can be implemented in accordance with the correct time and place of action. Closer inspection of the I Ching system (eg. as seen visually in the collaborative work "64s" in Chapter seven), however, reveals an underlying binary system - identical to that of the logical structure of if-then statements and indeed the very workings of the computer itself. As a system that can extend in a combinatorial fashion to any number of dimensions, at variable resolutions of scale, and that contains and expresses numerous structural relationships within itself and yet remains free from problematic linguistic associations, it is said that this is the theory that inspired Leibniz to derive his differential calculus [Baudrillard 1983 p103].

My works have included many such different control structures, often allowing that these in turn be alterable as a component of interactive play. Sonic control of the audio environment is thus a compositional element in this work, composing interactions by deciding which methods and interdependencies are best suited to each individual context for structuring a sound's production.

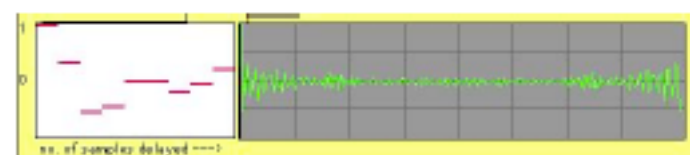
I will next outline the component listening models, fundamental as they are to any contextual control structure hoping to contain them. I will leave the overriding organizations largely unspecified, however, open for the imagination instead - it is after all better, perhaps, for such matters to be described in sound, not in word.

acf~





• PITCHED "ooo" at input
 on left - zoom in at start (first 9 samples)
 on right - whole vector buffer (512 samples)



• NOISY "sss" at input
 on left - zoom in at start (first 9 samples)
 on right - whole vector buffer (512 samples)

The acf~ model finds the autocorrelation function (acf) of an input signal. It is a time-domain analysis technique that is useful in beginning an assessment of the instantaneous timbre, the current sound quality, of an input signal. The method relies on a fact that Stan Tempelaars points out - **"the autocorrelation function of a periodic signal is itself also periodic with the same value of the period"** [Tempelaars 1996 p262], and a step-by-step way to create this signal is given in his book (pages 261-263).

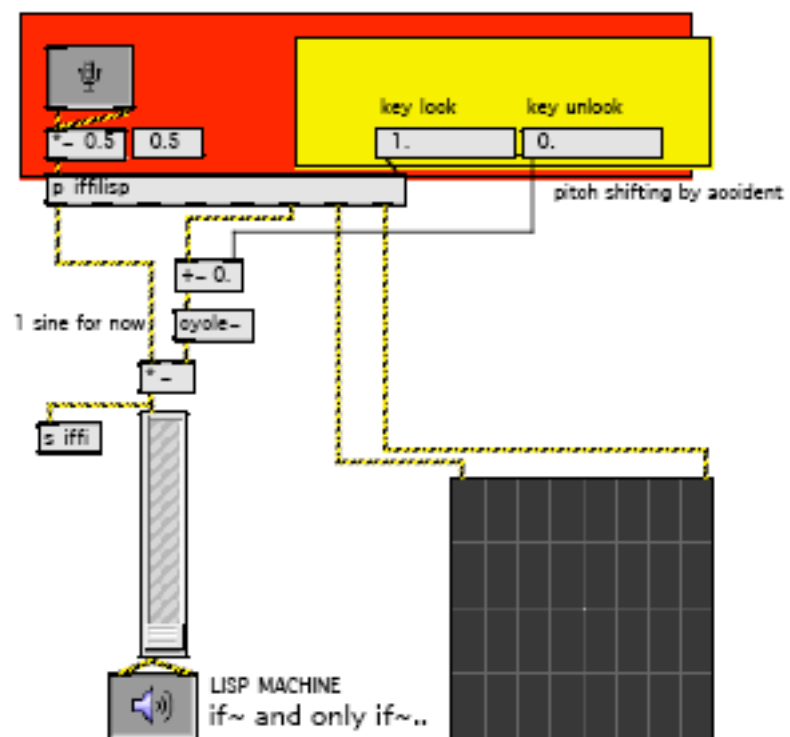
The autocorrelation method works by comparing an input signal with ever-so-slightly delayed copies of itself (1 sample delay, 2, 3, etc.), noting how this signal evolves with the number of samples delayed. When a signal is periodic these overlapped delayed copies of the input will add up to give a stronger peak in the acf output (more frequently for high pitched waves than for low). However, a noisy tone will not show these strong peaks.

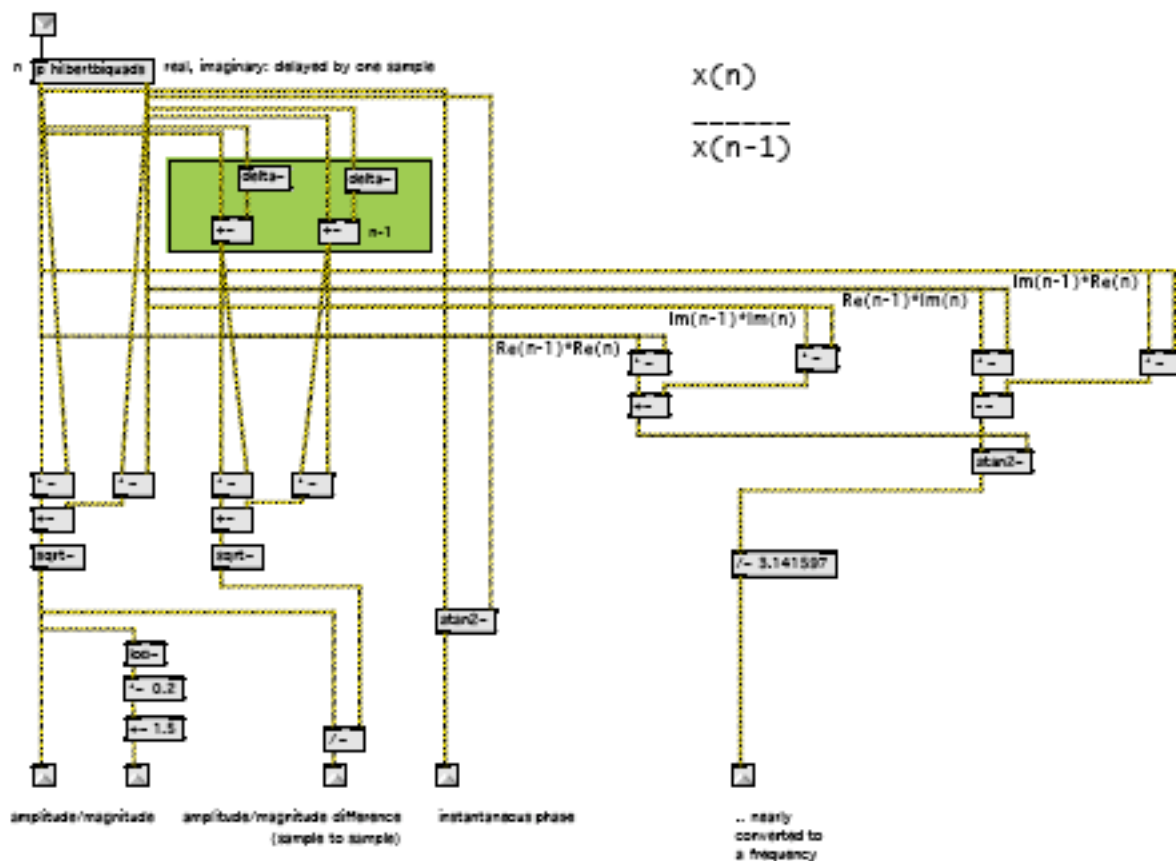
The acf can be re-interpreted to give, separately, a low frequency component describing the spectral envelope (the overall energy density distribution curve on the frequency axis) and higher frequency components representing the pitched aspects of the signal that apportion their energy throughout this spectral shape.

However, the factor of the model currently in use concerns the spectral slope around time zero, that is, the gradient apparent in the fall-off of the first few samples in each buffer. The first point in the acf buffer will always have the value one (by nature of the mathematical process used to create it) but pitch and noise will make very obvious differences to the next few samples. For a pitched input (a spoken "ooo"), a signal by nature repetitive in time with energy in regular vibrational patterns, the first nine samples (shown in the 'zoomed' section at left) are seen to all remain 'high'. However, for a noisy input (a spoken "sss") we see that the curve slopes steeply down in the first few samples delayed, the sound having already lost its correlation even over such a short delay period.

A simple gating mechanism on early sample values of this function provides a good pitch/noise switch. This mechanism provided the voice controller for "Ecce Space" (see Chapter seven), and is often used in conjunction with the reel8 model to match timbres or shades of noise.

if~





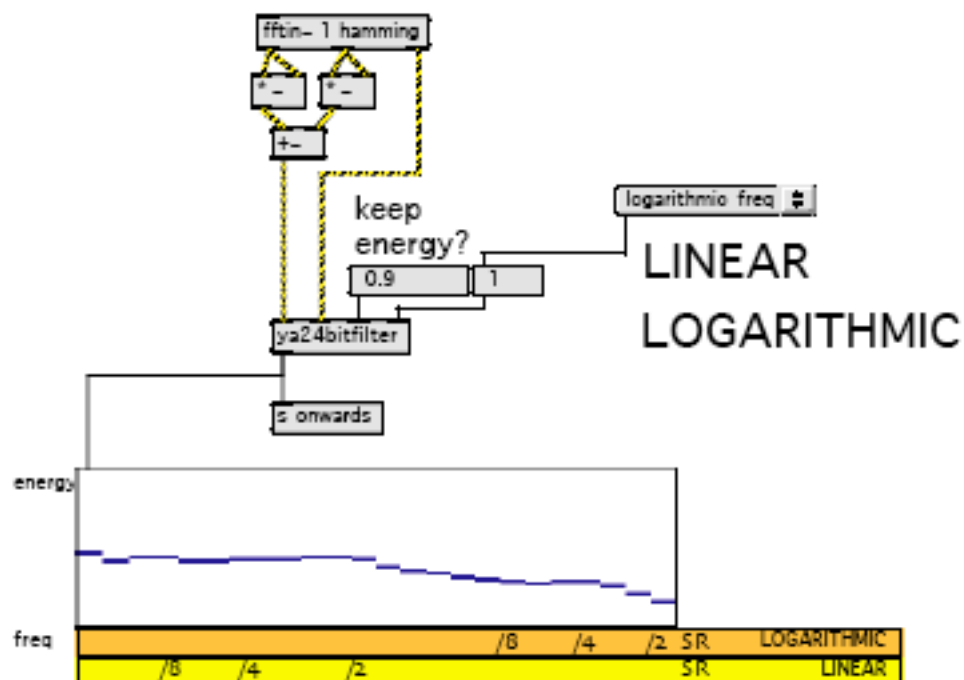
The if~ model, named for 'instantaneous frequency', is a time-domain technique by which to achieve an estimate of the frequency of the input signal on a moment by moment, sample by sample basis.

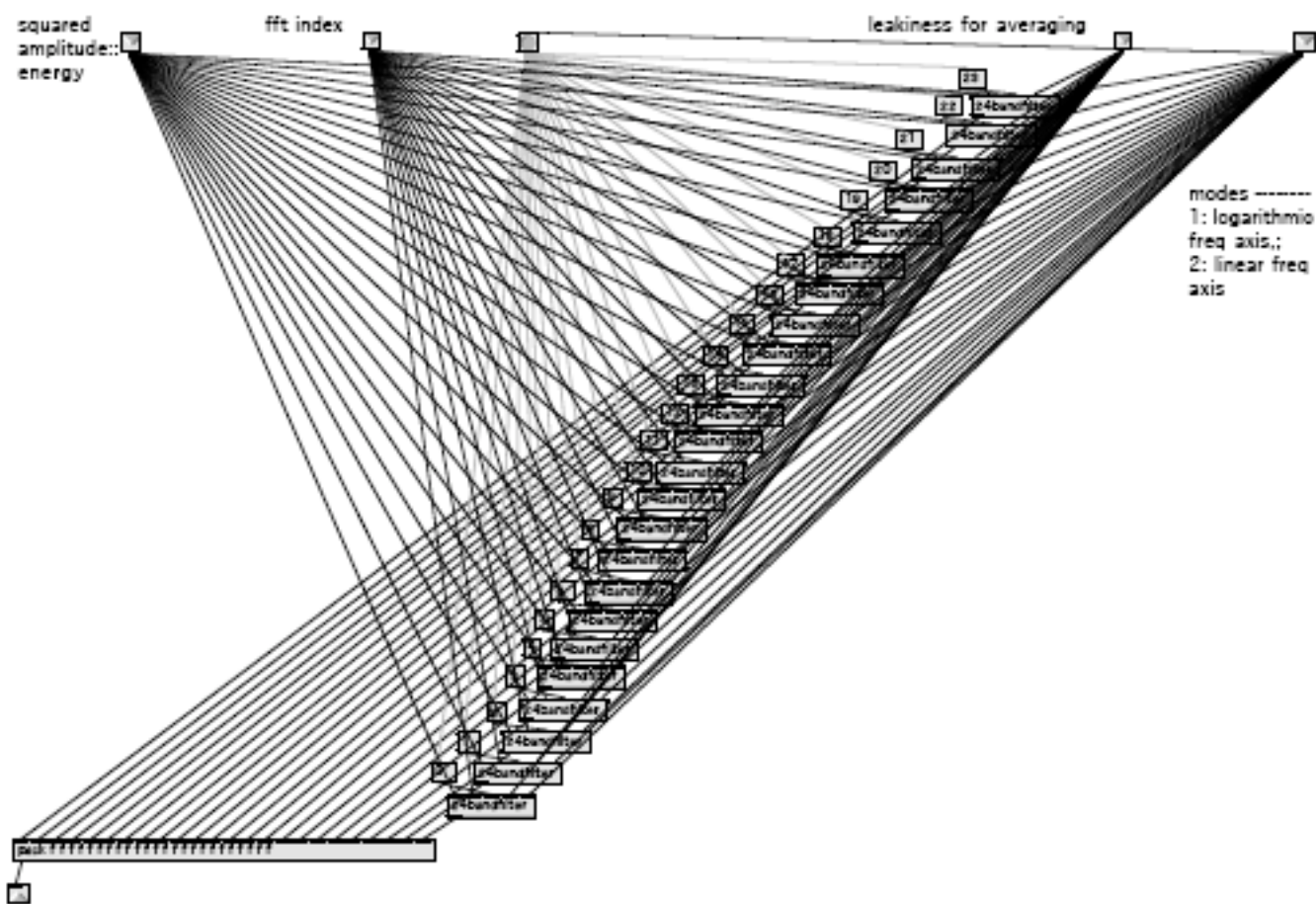
Any input to the if~ model is passed directly to a Hilbert analysis module whose job it is to create two signals at ninety degrees phase difference with each other. These are the real and imaginary parts of the signal, from which the amplitude and phase can be derived. The remainder of the model then performs a single complex division (ie. using imaginary number calculus), using the currnet $x(n)$ and previous $x(n-1)$ values of its input to determine, from one sample to the next, the changes that have occurred. The major benefit to this method of analysis is that it does not rely on buffering or windowing of any kind, and thus reports its results at audio rate, 'instantaneously'. The analysis is monophonic in that only one frequency result will be reported at a time, however the model runs at such a speed (reporting 44100 results per second for a typical sample rate) that it can still be useful to deal with input signals containing rich harmonic spectra.

A useful by-product of this analysis is the instantaneous frequency audio signal produced at output. Derived from the instantaneous phase-difference signal (a function which can easily be re-mapped back into audible ranges just with multiplication), it can be sampled to the control rate and used to manipulate any variable within further programs, or perhaps more interestingly, it can be to listened to. When this occurs we find that, for instance using a voice as input, frequencies reproduced at the output are limited (at around 6kHz or above) by the implementation of the hilbert analysis that has been used to produce the necessary analytical signals - real and imaginary, at 90 degrees phase difference. Thus for spoken words vowels are reproduced without distortion whereas the high-frequency transients (ss, sh, ff, etc.) are altered significantly. The result is that the if~ model re-synthesis, while sounding similar to the input, appears to lisp.

A further feature of the if~ model is that it turns out to be easy to perform transposition and 'harmonization' on such signals simply using the very basics of mathematical conditioning - addition and multiplication.

fsum~





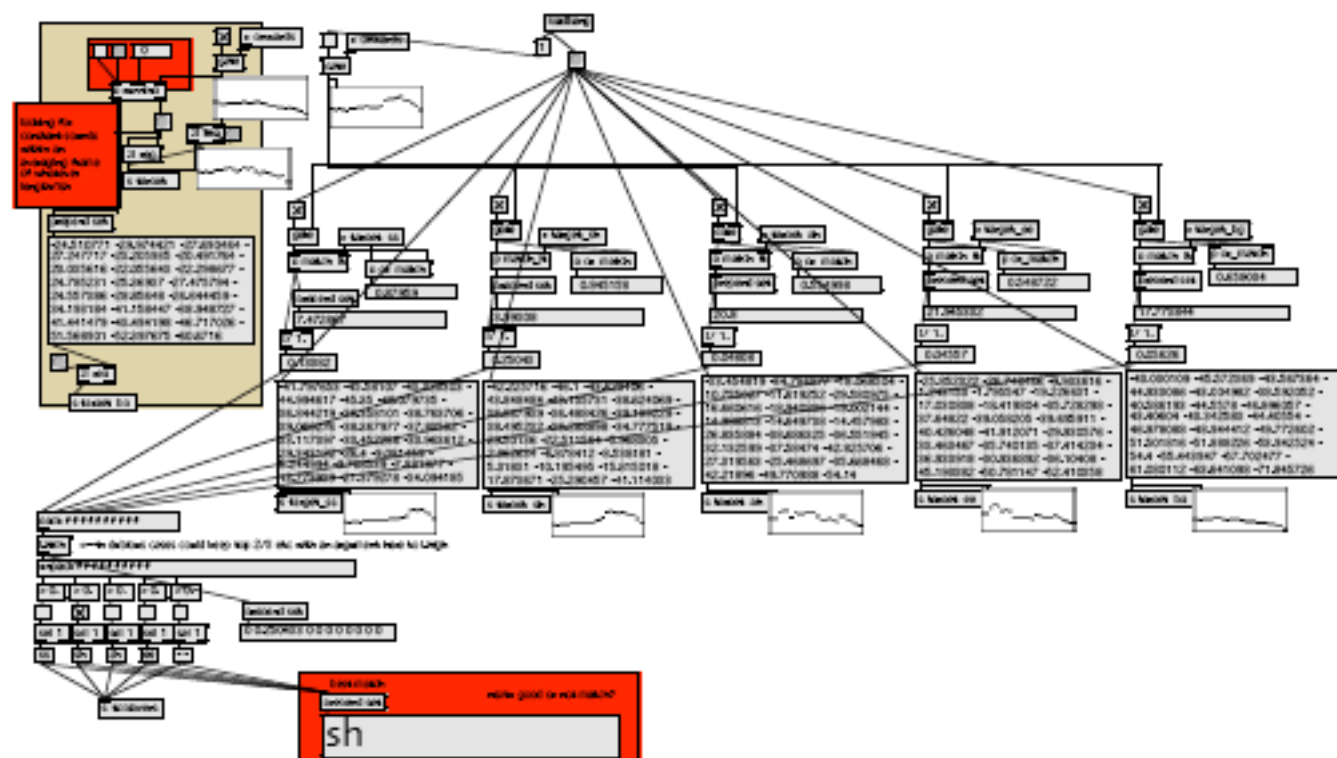
The fsum~ model allows the spectral content of an audio signal to be summarized as a list of numbers. Similar to the display given in a standard frequency analyzer, this model runs a Fourier analysis on the input signal to display (across a horizontal frequency axis) the energies present within the signal. The method is implemented with techniques of third-octave filter bank principles such that 24 individual filters, 3 per octave, cover eight octaves of sound.

The fsum~ model has two particular benefits. Firstly, alongside a traditional linear frequency arrangement (the one by which we think of white noise as having a 'flat' frequency spectrum), the frequency axis can also be scaled logarithmically. This 'points' the analysis towards the lower frequency areas in the spectrum, giving a higher sensitivity or better resolution on sounds which are closer to those we traditionally think of as pitches rather than noise, closer to those of the human voice. On this logarithmic scaling, more in line with our perception than is the linear scaling, white noise is seen in fact to have an increase in energy of 3dB per octave, and it is rather pink noise that displays the characteristically flat spectral shape.

The second benefit to this frequency analyzer is that as each filter acts independently to gather data on how much energy is present within its own particularly bounded area of reading, each filter can be 'smoothed' individually. Using the method described in Chapter eight, each filter is fitted internally with an integrator unit that averages its output between the current and previous values. This allows that the sensitivity of response (fast, slow) can be set variably across the frequency spectrum and thereby allows a more malleable data conditioning set to be implemented.

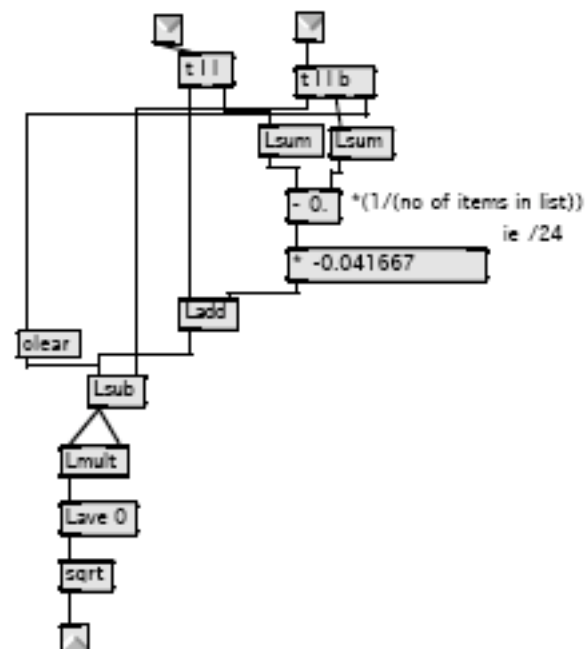
The fsum~ model is one of two methods discussed here that crosses the domain in MaxMsp from 'audio' to 'control' rates of work. Its output (a list of 24 numbers that update at control rate, once per window size in the FFT analysis stage) is commonly used as input to the reel8 model in order to perform spectral template pattern-matching tasks as were described in Chapter six in discussion of the voice-recognition patch "ssshahee--".

reel8~



incoming...

the one to find...



The reel8 model is a pattern-matching method that is implemented to perform timbral- and voice-recognition tasks throughout my work. It is based on finding the difference between an input signal and a particular target template (each represented as a list), and uses a mathematical formulation of 'standard deviation' in relation to a probabilistic spread of variables [Boas 1983 p709-710].

This unit was created initially to deal with the list results that are obtained from the fsum~ model though it works equally well with the lists resulting from the acf~ model. The modular nature of the process means that varying configurations can be used - multiple units placed side by side, for instance, and all driven with the same input signal, can determine which of a number of targets the input is closest to. This has been discussed in Chapter six in relation to the "ssshahee--" controller.

The incoming list (representing in this case the spectral properties of the sound, the energy distribution envelope) is received in the left input and compared with the pre-loaded target at the right inlet. The mathematical method dictates the sequence - each list is squared, value by value, and their difference is found. An extra step of normalization is included to discount the fact of a variable amplitude (or overall signal size) which would otherwise judge a loud sound to have a large difference from a quiet one of the same spectral distribution. The output of the unit reduces the list further to a single number which approaches zero for a very close match between input and target. Many other measures of difference are possible, and depend largely on how (in one's native tongue) one overcomes confusion and defines the concept of 'difference' - each reconfiguration points the sensitivity of the device into a slightly different area.

An interesting implementation using this technique also arisen while working with Vicky Lomas on an environment for (blokfluit) recorder and computer, "TTT" [2004]. This combines two such units, matching for different frequency centres in multiphonic tones produces by over-blowing the instrument, but combines the output as a single slider that moves between two values. Work also continues to 'slow down' the results of such comparisons - sequences of matched conditions can thus be found by creating longer data structures, summarized and re-strung together.

chapter ten

This thesis has outlined how, today, we work both on the physical, actualizing side of the art and technology, and simultaneously on the aesthetic, theorizing side of the art and technology. The former governs how the realization of a work will fit into its final context, and the latter describes the conceptual framework underlying the choices in the modeling by which sound is understood and produced.

Previous chapters have shown that communication between humans and machines most often deal with direct physical contacts and responses, and the varying degrees of sensitivity that these scales should be differentiated along. Our skills as humans have not changed significantly over recent millennia, but we now have the chance to open ourselves to the implications of recent technological developments and expand our minds thereby. Children, for instance, still struggle to catch a ball, whilst professional snooker players through training in the conventions of their 'play', hone skills that, while still based in the same hand-eye coordination tasks, allow concepts of reflective angles and indirect action to be explored. We live today knowing that **"certain dualisms have been persistent in western traditions ... High-tech culture challenges these dualisms in intriguing ways. It is not clear who makes and who is made in the relation between human and machine ... One consequence is that our sense of connection to our tools is heightened"** [Harraway 2000 p313].

Music is a response to the pattern and organization of life around us. Some artists make use of music's ability to transport us, to calm us down and escape the business outside of daily life, offering us a space in which to reflect - "careful listening is more important than making sounds happen" [Lucier 1995]. Some artists build further on top of such ideas and instead provide an auditory environment whose musical feel can be influenced live and is guided by the listening process. Such an environment uses aural/sonic methods of interaction, participation and response in order to encourage collaboration and partnership among people and with machines. A feeling of support, through the formation of groups that welcome the uninitiated participant, can build an audience by making them feel included as part of the event directly.

We are reminded that, in any case, **"music will continue to be produced ... but without the**

ferment of an actively engaged audience it will lapse into yet another form of consumerism" [Rowe 2001 p5]. Whatever 'happens' in an interactive audio installation, it should be the listening that guides the participant or performer to make their next action. Similarly, using computational listening models is of benefit - particularly to divide up the audio signal according to its particular qualities under investigation. Using such models as components in an installation setting allows the sound artist to compose interactions by observing and creating interdependencies amongst the individual components, encouraging a dynamic behaviour to emerge by **"describing modes of interaction and degrees of freedom within and between multiple agents" [Rowe 2001 p373].**

When working with sound in this manner, we find that **"the [sonic] interface design becomes then the very object of composition ... and the methods by which [the individual components] communicate among themselves should be seen as the material implementation of a compositional process or concept. This approach by which one invents and works out interdependencies among real-time control variables, already reflects a paradigm shift from interactive composing ... to composing interactions"** [Di Scipio 2003 p270]. Robert Rowe also reminds us that when components are considered as such, **"their behavior [can change] as a function of the musical context going on around them ... this versatility represents the essence of interaction and an intriguing expansion of the craft of composition"** [Rowe 2001 p4].

Aware that the nature of the interface itself (sonic or otherwise) is a prompt to the player to adopt a certain behaviour pattern, some artists explore the resulting implications by combining a wide range of different interaction devices in one time in the same physical space [eg. Ulyate and Bianciardi 2002]. Such interactive works can allow for control of audio in terms of both the actual content and its presentation in the room. By building environments flexible enough to capture and maintain interest and attention in both the musician and non-musician, such artists encourage listening as a pre-requisite for conscious thought about the interface, challenging people to consider and to become aware of how they are interacting with the sound present.

Installations have been shown to be fertile ground for an exploration of the issue of live input-output relations, to do with the live 'now' of a musical moment. Freed from some of the cultural constraints that traditionally occupy a music's composition and performance, installation artworks can provide a context in which both of these elements may be

easily included, questioning the relationship between live and pre-composed aspects of a work. By placing the audience into a completely new relationship with the stage, interactive installations promote cooperation between people, and force us to become more focused in finding concepts that allow us to share our intentions and experiences.

"The concept of an interactive instrument may, in the near future, define a new way for the public to experience music" [Chadabe 2002 p2].

references

Arfib, D. and Kessous, L. [2003] Gestural Control of Sound Synthesis and Processing Algorithms, chapter in Gesture and Sign Language in Human-Computer Interaction: International Gesture Workshop, GW 2001, London pp 285-295

Baudrillard, J. [1983] Simulations, Semiotext[e] and Jean Baudrillard, USA

Bell, D. and Kennedy, B.M. -eds. [2000] The Cybercultures Reader, Routledge, London and New York

Boas, M.L. [1983] Mathematical Methods in the Physical Sciences, First Edition 1966, Second Edition 1983, John Wiley & Sons, US and Canada

Bolter, J.D. and Gruslin, R. -eds. [2000] Remediation: Understanding New Media, The MIT Press, Massachusetts and London

Camurri, A., Trocca, R. and Volpe, G. [2002] Interactive Systems Design: A KANSEI-based Approach
<http://hct.ece.ubc.ca/nime/2002/proceedings/paper/camurri.pdf>

Chadabe, J. [2002] The Limitations of Mapping as a Structural Descriptive in Electronic Instruments, Proceedings NIME-02, Dublin

Choi, I. [1997] A Chaotic Oscillator as a Musical Signal Generator in an Interactive Performance System, Journal of New Music Research, Vol 26, pp 17-47

Cox, C. and Warner, D. - Eds. [2004] Audio Culture: Readings in Modern Music, Continuum, New York and London

Daubner, E. [1999] Interactive Strategies and Dialogical Allegories: Encountering David Rokeby's Transforming Mirrors Through Marcel Duchamp's Open Windows and Closed Doors, presented at: "Invencão: Thinking the Next Millenium", São Paulo, Brazil
<http://www.itaucultural.org.br/invencao/papers/Daubner.htm>
<http://homepage.mac.com/davidrokeby/daubner.html>

Di Scipio, A. [2003] 'Sound is the interface': from interactive to ecosystemic signal processing, Organised Sound 8(3): pp 269-277

Durant, A. [1990] A new day for music?, essay contained in Culture, Technology and Creativity pp 175-196 - see entry under Hayward

Endelt, L.O. and la Cour-Harbo, A. [2004] Comparison of Methods for Sparse Representation of Musical Signals
<http://www.control.auc.dk/~alc/art-comparison-ICASSP-2005.pdf>

Freed, D. J. [1990] Auditory correlates of perceived mallet hardness for a set of recorded percussive sound events, Journal of the Acoustical Society of America, vol. 87, pp 311-322

Grey, J.M. and Gordon, J.W. [1978] Perceptual effects of spectral modifications on musical timbres, Journal of the Acoustical Society of America, vol. 63(5), pp 1493-1500

Harraway, D. [2000] A Cyborg Manifesto, essay contained in the Cybercultures Reader - see Cybercultures Reader entry under Bell.

Hayward, P. -ed [1990] Culture Technology and Creativity in the late twentieth century. John Libbey & Company Ltd., London, Montrouge and Rome.

Hirata, K. and Matsuda, S. [2002] Interactive Music Summarization based on GTTM
ismir2002.ismir.net/proceedings/02-FP03-2.pdf

Hunt, A., Wanderley, M.M. and Paradis, M. [2002] The importance of parameter mapping in electronic instrument design, Proceedings NIME-02, Dublin

Igarashi, T. [2001] Grunts to control computers, news item reported by the BBC
news.bbc.co.uk/1/hi/english/sci/tech/newsid_1606000/1606175.stm

Jehan, T. [2005] Creating Music by Listening, PhD thesis submitted to MIT
<http://web.media.mit.edu/~tristan/phd/dissertation/>

Kahn, D. [1999] Noise Water Meat: A History of Sound in the Arts, The MIT Press, Massachusetts and London

Levitin, D.J., McAdams, S. and Adams, R.L. [2002] Control Parameters for musical instruments: a foundation for new mappings of gesture to sound, Organised Sound 7(2),

Lu, L. and Zhang, H.-J. [2003] Automated Extraction of Music Snippets
http://research.microsoft.com/users/llu/Publications/ACMMM03_MusicSnippet.pdf

Lucier, A. [1995] [Alvin Lucier, in *Reflexions: Interviews, Scores, Writings* (Köln: Edition MusikTexte, 1995) p430] - included in *Audio Culture* p63, see entry under Cox

(*) **Malik, H.** [2004] Content-Based Audio Indexing and Retrieval: an overview.
[CBAIR_survey.pdf, access date 4 December 2004]

Malik, K. [2001] What is it to be human? what science can and cannot tell us. Institute of Ideas, Conversations in Print. Academy of Ideas, London.

Mott, I. [2005] Summoned Voices
<http://www.reverberant.com>

Mandelis, J. and Husbands, P. [2003] Musical Interaction with Artificial Life Forms: Sound Synthesis and Performance Mappings, *Contemporary Music Review*, Vol 22 No 3, pp 69-77

Miller, J. and Carterette, P. [1975] Perceptual space for musical structures, *Journal of the Acoustical Society of America*, vol. 58(3), pp 771-720

Ritter, D. [1996] *The Intersection of Art and Interactivity*, originally published in *Ars Electronica Festival Memesis*, Springer, Vienna
<http://www.soundtoys.net/a/journal/texts/ritter.html>

Robson, D. [2002] PLAY!: Sound Toys for Non-Musicians, *Computer Music Journal*, 26: 3, pp 50-61

Robinson, D.J.M. [2002] Perceptual model for assessment of coded audio, PhD thesis submitted to the University of Essex.
http://web.media.mit.edu/~tristan/Courses/MAS.945/Papers/Technical/Robinson_Thesis.pdf

Rokeby, D. [1995] *Transforming Mirrors: Subjectivity and Control in Interactive Media*,

essay in Critical Issues in Electronic Media, ed. Simon Penny pp 133-158, State University of New York Press, New York

Rowe, R. [2001] Machine Musicianship. The MIT Press, Massachusetts and London

Ryan, J. [1991] Some Remarks on Musical Instrument Design at STEIM, Contemporary Music Review, 6(1):3-17

Sapir, S. [2002] Gestural Control of Digital Audio Environments, Journal of New Music Research, 31(2), pp 119-129

Scheirer, E. D. [2000] Music-Listening Systems. PhD thesis submitted to MIT
<http://sound.media.mit.edu/~eds/thesis>

Smaragdis, P. [2001] Redundancy Reduction for Computational Audition, A Unifying Approach. PhD thesis submitted to MIT
<http://web.media.mit.edu/~paris/phd/paris-cover.pdf>

Smith, J.D. [1997] The Place of musical novices in music science. Music Perception 14(3), pp 227-262

Song, J., Bae, S.Y. and Yoon, K. [2002] Mid-Level Music Melody Representation of Polyphonic Audio for Query-by-Humming System
<http://ismir2002.ismir.net/proceedings/02-FP04-5.pdf>

Spevak, C. and Favreau, E. [2002] Soundspotter - A prototype System for Content-based Audio Retrieval
<http://www2.hsu-hh.de/ant/dafx2002/papers.html>

Stuart, C. [2003] The Object of Performance: Aural Performativity in Contemporary Laptop Music, Contemporary Music Review, Vol 22 No 4, pp 59-65

Straus, J.N. [2004] Normalizing the Abnormal: Disability in Music and Music Theory
<http://web.gc.cuny.edu/Music/disability/straus.pdf>

Tempelaars, S. [1996] Signal Processing, Speech and Music, Swets & Zeitlinger B.V., Lisse

Ulyate, R. and Bianciardi, D. [2002] The Interactive Dance Club: Avoiding Chaos in a Multi-Participant Environment, *Computer Music Journal*, 26:3, pp 40-49

von Bismarck, G. [1974] Sharpness As an Attribute of Steady Sounds, *Acustica*, vol. 30, pp 156-159

Wanderley, M.M. and Orio, N. [2002] Evaluation of Input Devices for Musical Expression: Borrowing Tools from HCI, *Computer Music Journal*, 26:3, pp 62-76

Warwick, K. [2001] A cyborg life for me, contained in What is it to be human? pp 42-44 - see entry under Malik, K.

Wessel, D. and Wright, M. [2002] Problems and Prospects for Intimate Musical Control of Computers, *Computer Music Journal*, 26:3, pp 11-22

Whitelaw, M. [2003] Sound Particles and Microsonic Materialism, *Contemporary Music Review*, Vol 22 No 4, pp 93-100

Wiener, N. [1961] *Cybernetics: Or, Control and Communication in the Animal and in the Machine*, The MIT Press, Cambridge. 2nd edition 1961, 1st edition 1948

Wilson, S. [2002] *Information Arts: Intersections of Art, Science, and Technology*, The MIT Press, Massachusetts and London

Winkler, T. [1995] *Strategies for Interaction: Computer Music, Performance, and Multimedia*, Proceedings of the 1995 Connecticut College Symposium on Arts and Technology

Wishart, T. [2002] *On Sonic Art* (First Edition 1996, New and Revised edition, 2002) Routledge, London

Woolf, S. and Yee-King, M. [2003] *Virtual and Physical Interfaces for Collaborative Evolution of Sound*

All web-based materials were accessed during the period between late 2003 and early 2005 and all bar one (*) are currently, as of 8 September 2005, held online at the URLs given here