AUDIBLE DESIGN

A PLAIN AND EASY INTRODUCTION TO PRACTICAL SOUND COMPOSITION

by

TREVOR WISHART

Published by Orpheus the Pantomime Ltd : 1994

PREFACE

The main body of this book was written at the Institute of Sonology at the Royal Conservatory in the Hague, where I was invited as composer in residence by Clarence Barlow, in 1993. Some clarifications and supplementary material was added after discussions with Miller Puckette, Zak Settel, Stephan Bilbao and Philippe Depalle at IRCAM. However, the blame for any inaccuracies or inconsistencies in the exposition rests entirely with me. The text of the book was originally written entirely in longhand and I am indebted to Wendy McGreavy for transferring these personal pretechnological hieroglyphs to computer files. Help with the final text layout was provided by Tony Myatt of the University of York.

My career in music-making with computers would not have been possible without the existence of a community of like-minded individuals committed to making powerful and open music-computing tools available to composers on affordable domestic technology. I am therefore especially indebted to the *Composers Desktop Project* and would hence like to thank my fellow contributors to this long running project; in particular Tom Endrich, the real driving force at the heart of the CDP, to whom we owe the survival, expansion and development of this cooperative venture, against all odds, and whose persistent probing and questioning has led to clear instrument descriptions & musician-friendly documentation; Richard Orton and Andrew Bentley, the other composer founder members of, and core instrument contributors to the project; David Malham who devised the hardware bases of the CDP and much else, and continues to give support; Martin Atkins, whose computer science knowledge and continuing commitment made, and continues to make, the whole project possible (and from whom I have very slowly learnt to program less anarchically, if not yet elegantly!); Rajmil Fischman of the University of Keele, who has been principally responsible for developing the various graphic interfaces to the system; and to Michael Clarke, Nick Laviers, Rob Waring, Richard Dobson and to the many students at the Universities of York, Keele, and Birmingham and to individual users elsewhere, who have supported, used and helped sustain and develop this resource.

All the sound examples accompanying this book were either made specifically for this publication, or come from my own compositions *Red Bird*, The *VOX Cycle* or *Tongues of Fire*, except for one item, and I would like to thank Paul de Marinis and Lovely Music for permitting me to use the example in Chapter 2 from the piece *Odd Evening* on the CD *Music as a Second Language* (Lovely Music LCD 3011). Thanks are also due to Francis Newton, for assistance with data transfer to DAT.

CONTENTS

CHAPTER 0

INTRODUCTION

WHAT THIS BOOK IS ABOUT

This is a book about composing with sounds. It is based on three assumptions.

1. Any sound whatsoever may be the starting material for a musical composition.

2. The ways in which this sound may be transformed are limited only by the imagination of the composer.

3. Musical structure depends on establishing audible relationships amongst sound materials.

The first assumption can be justified with reference to both aesthetic and technological developments in the Twentieth Century. Before 1920, the French composer Varese was imagining a then unattainable music which had the same degree of control over sonic substance as musicians have traditionally exercised over melody, harmony and duration. This concern grew directly out of the sophisticated development of orchestration in the late Nineteenth Century and its intrinsic limitations (a small finite set of musical instruments). The American composer John Cage was the first to declare that all sound was (already) music. It was the emergence and increasing perfection of the technology of sound recording which made this dream accessible.

The exploration of the new sounds made available by recording technology was begun by Pierre Schaeffer and the G.R.M. in Paris in the early 1950s. Initially hampered by unsophisticated tools (in the early days, editing between lacquer discs – later the transformations – like *tape speed variation*, *editing* and *mixing* – available with magnetic tape) masterpieces of this new medium began to emerge and an approach to musical composition rooted in the sound phenomenon itself was laid out in great detail by the French school.

The second of our assumptions had to await the arrival of musical instruments which could handle in a subtle and sophisticated way, the inner substance of sounds themselves. The digital computer provided the medium in which these tools could be developed. Computers allow us to digitally record any sound at all and to digitally process those recorded sounds in any way that we care to define.

In this book we will discuss in general the properties of different kinds of sound materials and the effects certain well defined processes of transformation may have on these. We will also present, in the Appendix, a simple diagrammatic explanation of the musical procedures discussed.

The third assumption will either appear obvious, or deeply controversial, depending on the musical perspective of the reader. For the moment we will assume that it is obvious. The main body of this book will therefore show how, starting from a given sound, many other audibly similar sounds may be developed which however, possess properties different or even excluded from the original sound. The question of how these relationships may be developed to establish larger scale musical structures will be suggested towards the end of the book but in less detail as, to date, no universal tradition of large scale form-building (through these newly accessible sound-relationships) has established itself as a norm.

WHAT THIS BOOK IS NOT ABOUT

This book is *not* about the merits of computers or particular programming packages. However, most of the programs described were available on the *Composers Desktop Project* (CDP) System at the time of writing. The CDP was developed as a composers cooperative and originated in York, U.K.

Nor will we, outside this chapter, discuss whether or not any of the processes described can be, or ought to be implemented in real time. In due course, many of them will run in real time environments. My concern here, however, is to uncover the musical possibilities and restraints offered by the medium of sonic composition, not to argue the pros and cons of different technological situations.

A common approach to sound–composition is to define "instruments" – either by manipulating factory patches on a commercial synthesizer, or by recording sounds on a sampler – and then trigger and transpose these sounds from a MIDI keyboard (or some other kind of MIDI controller). Many composers are either forced into this approach, or do not see beyond it, because cheaply available technology is based in the note/instrument conception of music. At its simplest such an approach is no more than traditional note–oriented composition for electronic instruments, particularly where the MIDI interface confines the user to the tempered scale. This is not significantly different from traditional on–paper composition and although this book should give some insight into the design of the 'instruments' used in such an approach, I will not discuss the approach as such here – the entire history of European compositional theory is already available!

On the contrary, the assumption in this book is that we are not confined to defining "instruments" to arrange on some preordained pitch/rhythmic structure (though we may choose to adopt this approach in particular circumstances) but may explore the multidimensional space of sound itself, which may be moulded like a sculptural medium in any way we wish.

We also do not aim to cover every possibility (this would, in any case, be impossible) but only a wide and, hopefully, fairly representative, set of processes which are already familiar. In particular, we will focus on the transformation of sound materials taken from the real world, rather than on an approach through synthesis. However, synthesis and analysis have, by now, become so sophisticated, that this distinction need barely concern us any more. It is perfectly possible to use the analysis of a recorded sound to build a synthesis model which generates the original sound and a host of other related sounds. It is also possible to use sound transformation techniques to change any sound into any other via some well–defined and audible series of steps. The common language is one of intelligent and sophisticated sound transformation so that sound composition has become a plastic art like sculpture. It is with this that we will be concerned.

THINKING ALOUD – A NEW CRITICAL TRADITION

I cannot emphasise strongly enough that my concern is with the world of sound itself, as opposed to the world of notations of sound, or the largely literary disciplines of music score analysis and criticism. I will focus on what can be aurally perceived, on my direct response to these perceptions and on what can be technically, acoustically or mathematically described.

The world of sound–composition has been hampered by being cast in the role of a poor relation to more traditional musical practice. In particular the vast body of analytical and critical writings in the musicology of Western Art music is strongly oriented to the study of musical texts (scores) rather than to a discipline of acute aural awareness in itself. Sound composition requires the development of both new listening and awareness skills for the composer and, I would suggest, a new analytical and critical discipline founded in the study of the sonic experience itself, rather than its representation in a text. This new paradigm is beginning to struggle into existence against the immense inertia of received wisdom about 'musical structure'.

I have discussed elsewhere (*On Sonic Art*) the strong influence of mediaeval 'text worship' on the critical/analytical disciplines which have evolved in music. Both the scientific method and technologised industrial society have had to struggle against the passive authority of texts declaring eternal truths and values inimical to the scientific method and to technological advance.

I don't wish here to decry the idea that there may be "universal truths" about human behaviour and human social interaction which science and technology are powerless to alter. But because our prime medium for the propagation of knowledge is the written text, powerful institutions have grown up around the presentation, analysis and evaluation of texts and textual evidence .... so powerful that their influence can be inappropriate.

In attempting to explore the area of composing with sound, this book will adopt the point of view of a scientific researcher delving into an unknown realm. We are looking for evidence to back up any hypotheses we may have about potential musical structure and this evidence comes from our perception of sounds themselves. (Scientific readers may be surprised to hear that this stance may be regarded as polemical by many musical theorists).

In line with this view, therefore, this book is not intended to be read without listening to the sound examples which accompany it. In the scientific spirit, these are presented as evidence of the propositions being presented. You are at liberty to affirm or deny what I have to say through your own experience, but this book is based on the assumption that the existence of structure in music is a matter of fact to be decided by listening to the sounds presented, not a matter of opinion to be decided on the authority of a musical text (a score or a book .. even this one), or the importance of the scholar or composer who declares structure to be present (I shall return to such matters in particular in Chapter 9 on 'Time').

However, this is not a scientific text. Our interest in exploring this new area is not to discover universal laws of perception but to suggest what might be fruitful approaches for artists who wish to explore the vast domain of new sonic possibilities opened up by sound recording and computer technology.

We might in fact argue that the truly potent texts of our times are certainly not texts like this book, or even true scientific theories, but computer programs themselves. Here, the religious or mystical potency with which the medieval text was imbued has been replaced by actual physical efficacy. For the text of a computer program can act on the world through associated electronic and mechanical hardware, to make the world anew, and in particular to create new and unheard sonic experiences. Such texts are potent but, at the same time, judgeable. They do not radiate some mystical authority to which we must kow–tow, but do something specific in the world which we can judge to be more or less successful. And if we are dissatisfied, the text can be modified to produce a more satisfactory result.

The practical implications of this are immense for the composer. In the past I might spend many days working from an original sound source subjecting it to many processes before arriving at a satisfactory result. As a record of this process (as well as a safeguard against losing my hard-won final product) I would keep copious notes and copies of many of the intermediate sounds, to make reconstruction (or variation) of the final sound possible. Today, I store only the source and a brief so-called "batch-file". The text in the batch-file, if activated, will automatically run all the programs necessary to create the goal sound. It can also be copied and modified to produce whatever variants are required.

An illuminating manuscript indeed!

SOUND TRANSFORMATION : SCIENCE OR ART ?

In this book we will refer to sounds as sound-materials or sound-sources and to the process of changing them as transformations or sound-transformations. The tools which effect changes will be described as musical instruments or musical tools. From an artistic point of view it is important to stress the continuity of this work to past musical craft. The musical world is generally conservative and denizens of this community can be quick to dismiss the "new-fangled" as unmusical or artistically inappropriate. However, we would stress that this is a book by, and for, musicians.

Nevertheless, scientific readers will be more familiar with the terms signal, signal processing and computer program or algorithm. In many respects, what we will be discussing is signal processing as it applies to sound signals. However, the motivation of our discussion is somewhat different from that of the scientific or technological study of signals. In analysing and transforming signals for scientific purposes, we normally have some distinct goal in mind - an accurate representation of a given sequence, extraction of data in the *frequency domain*, the removal of noise and the enhancement of the signal "image" - and we may test the result of our process against the desired outcome and hence assess the validity of our procedure.

In some cases, musicians share these goals. Precise analysis of sounds, extraction of time-varying information in the frequency domain (see **Appendix p3**), sound clarification or noise reduction, are all of great importance to the sonic composer. But beyond this, the question that a composer asks is, is this process aesthetically interesting? - does the sound resulting from the process relate perceptually and in a musically useful manner to the sound we began with? What we are searching for is a way to transform sound material to give resulting sounds which are clearly close relatives of the source, but also clearly different. Ideally we require a way to "measure" or order these degrees of difference allowing us to articulate the space of sound possibilities in a structured and meaningful way.

Beyond this, there are no intrinsic restrictions on what we do. In particular, the goal of the process we set in motion may not be known or even (with complex signals) easily predictable beforehand. In fact, as musicians, we do not need to "know" completely what we are doing (!!). The success of our efforts will be judged by what we hear. For example, a technological or scientific task may involve the design of a highly specific *filter* to achieve a particular result. A musician, however, is more likely to require an extremely flexible (band-variable, Q-variable, time-variable: **Appendix p8**) filter in order to *explore* its effects on sound materials. He/she may not know beforehand exactly what is being searched for when it is used, apart from a useful aesthetic transformation of the original source. What this means in practice may only emerge in the course of the exploration.

This open-ended approach applies equally to the design of musical instruments (signal processing programs) themselves. In this case, however, it clearly helps to have a modicum of acoustic knowledge. An arbitrary, number-crunching program will produce an arbitrary result - like the old adage of trying to write a play by letting a chimpanzee stab away at a typewriter in the hope that a masterpiece will emerge.

So scientists be forewarned! We may embark on signal-processing procedures which will appear bizarre to the scientifically sophisticated, procedures that give relatively unpredictable results, or that are heavily dependent on the unique properties of the particular signals to which they are applied. The question we must ask as musicians however is not, are these procedures scientifically valid, or even predictable, but rather, do they produce aesthetically useful results on at least some types of sound materials.

There is also a word of caution for the composer reader. Much late Twentieth Century Western art music has been dogged by an obsession with complicatedness. This has arisen partly from the permutational procedures of late serialism and also from an intellectually suspect linkage of crude information theory with musical communication - more patterns means more information, means more musical "potency". This obsession with quantity, or information overload, arises partly from the breakdown of consensus on the substance of musical meaning. In the end, the source of musical potency remains as elusive as ever but in an age which demands everything be quantifiable and measurable, a model which stresses the *quantity* of information, or *complicatedness* of an artefact, seems falsely plausible.

This danger of overkill is particularly acute with the computer- processing of sound as anything and everything can be done. For example, when composing, I may decide I need to do something with a sound which is difficult, or impossible, with my existing musical tools. I will therefore make a new instrument (a program) to achieve the result I want. Whilst building the instrument, however, I will make it as general purpose as possible so that it applies to all possible situations and so that all variables can vary in all possible ways. Given the power of the computer, it would be wasteful of time not to do this. This does not mean however that I will, or even intend to, use every conceivable option the new instrument offers. Just as with the traditional acoustic instrument, the task is to use it, to play it, well. In sound composition, this means to use the new tools in a way appropriate to the sound we are immediately dealing with and with a view to particular aesthetic objectives. There is no inherent virtue in doing everything.

EXPLICIT AND INTUITIVE KNOWLEDGE

In musical creation we can distinguish two quite distinct modes of knowing and acting. In the first, a physical movement causes an immediate result which is monitored immediately by the ears and this feedback is used to modify the action. Learned through physical practice and emulation of others and aided by discussion and description of what it involved, this real-time-monitored action type of knowledge, I will describe as *intuitive*". It applies to things we know very well (like how to walk, or how to construct meaningful sentences in our native tongue) without necessarily being able to describe explicitly what we do, or why it works. In music, intuitive knowledge is most strongly associated with musical performance.

On the other hand, we also have *explicit* knowledge of, e.g. acceptable HArmonic progressions within a given style or the spectral contour of a particular vowel formant (see Chapter 3). This is knowledge that we know that we know and we can give an explicit description of it to others in language or mathematics. Explicit knowledge of this kind is stressed in the training, and usually in the practice of, traditional composers.

In traditional "on paper" composition, the aura of explicitness is enhanced because it results in a definitive text (the score) which can often be given a rational exegesis by the composer or by a music score-analyst. Some composers (particularly the presently devalued "romantic" composers), may use the musical score merely as a notepad for their intuitive musical outpourings. Others may occasionally do the same but in a cultural atmosphere where explicit rational decision-making is most highly prized, will claim, post-hoc, to have worked it all out in an entirely explicit and rational way.

Moreover, in the music-cultural atmosphere of the late Twentieth Century, it appears "natural" to assume that the use of the computer will favour a totally explicit approach to composition. At some level, the computer must have an exact description of what it is required to do, suggesting therefore that the composer must also have a clearly explicit description of the task. (I will describe later why this is not so).

The absurd and misguided rationalist nightmare of the totally explicit construction of all aspects of a musical work is not the "natural" outcome of the use of computer technology in musical composition – just one of the less interesting possibilities.

It might be supposed that the pure electro-acoustic composer making a composition directly onto a recording medium has already abandoned intuitive knowledge by eliminating the performer and his/her role. The fact, however, is that most electro-acoustic composers 'play' with the medium, exploring through informed, and often real-time, play the range of possibilities available during the course of composing a work. Not only is this desirable. In a medium where *everything* is possible, it is an essential part of the compositional process. In this case a symbiosis is taking place between composerly and performerly activities and as the activity of composer and performer begin to overlap, the role of intuitive and explicit knowledge in musical composition, must achieve a new equilibrium.

In fact, in all musical practice, some balance must be struck between what is created explicitly and what is created intuitively. In composed Western art music where a musical score is made, the composer takes responsibility for the organisation of certain well-controlled parameters (pitch, duration, instrument type) *up to a certain degree of resolution*. Beyond this, performance practice tradition and the player's intuitive control of the instrumental medium takes over in pitch definition (especially in processes of succession from pitch to pitch on many instruments, and with the human voice), timing precision or its interpretation, and sound production and articulation.

At the other pole, the free-improvising performer relies on an intuitive creative process (restrained by the intrinsic sound/pitch limitations of a sound source, e.g. a retuned piano, a metal sheet). to generate both the moment-to-moment articulation of events and the ongoing time structure at all levels.

However, even in the free-improvisation case, the instrument builder (or, by accident, the found-object manufacturer) is providing a framework of restrictions (the sound world, the pitch-set, the articulation possibilities) which bound the musical universe which the free improviser may explore. In the sense that an instrument builder sets explicit limits to the performer's free exploration she/he has a restricting role similar to that of the composer.

In contrast, the role of computer instrument builder is somewhat different. Information Technology allows us to build sound-processing tools of immense generality and flexibility (though one might not guess this fact by surveying the musical hardware on sale in high street shops). Much more responsibility is therefore placed on the composer to choose an appropriate (set of) configuration(s) for a particular purpose. The "instrument" is no longer a definable (if subtle) closed universe but a groundswell of possibilities out of which the sonic composer must delimit some aesthetically valid universe. Without some kind of intuitive understanding of the universe of sounds, the problem of choice is insurmountable (unless one replaces it entirely by non- choice strategies, dice-throwing procedures etc).

REAL TIME OR NOT REAL TIME ? - THAT IS THE QUESTION

As computers become faster and faster, and more and more powerful, there is a clamour among musicians working in the medium for "real-time" system, i.e. systems on which we hear the results of our decisions as we take them. A traditional musical instrument is a "real-time system". When we bow a note on a violin, we immediately hear a sound being produced which is the result of our decision to use a particular finger position and bow pressure and which responds immediately to our subtle manual articulation of these. Success in performance also depends on a foreknowledge of what our actions will precipitate. Hence the rationale for having real-time systems seems quite clear in the sphere of musical performance. To develop new, or extended musical performance instruments (including real-time-controllable sound processing devices) we need real-time processing of sounds.

Composition on the other hand would seem, at first glance, to be an intrinsically non-real-time process in which considered and explicit choices are made and used to prepare a musical text (a score), or (in the studio) to put together a work, sound-by-sound, onto a recording medium out of real time. As this book is primarily about composition, we must ask what bearing the development of real-time processing has on compositional practice apart from speeding up some of the more mundane tasks involved.

Although the three traditional roles performer-improviser, instrument-builder and composer are being blurred by the new technological developments, they provide useful poles around which we may assess the value of what we are doing.

I would suggest that there are two conflicting paradigms competing for the attention of the sonic composer. The first is an instrumental paradigm, where the composer provides electronic extensions to a traditional instrumental performance. This approach is intrinsically "real time". The advantages of this paradigm are those of "liveness" (the theatre versus the cinema, the work is recreated 'before your very eyes') and of mutability dependent on each performer's reinterpretation of the work. This approach fits well into a traditional musical way of thinking.

It's disadvantages are not perhaps immediately obvious. But the composer who specifies a network of electronic processing devices around a traditional instrumental performance must recognise that he/she is in fact creating a new and different instrument for the instrumentalist (or instrumentalist-"technician" duo) to play and is partly adopting the role of an instrument builder with its own very different responsibilities. In the neomanic cultural atmosphere of the late Twentieth Century, the temptation for anyone labelled "composer" is to build a new electronic extension for every piece, to establish credentials as an "original" artist. However, an instrument builder must demonstrate the viability and efficacy of any new instrument being presented. Does it provide a satisfying balance of restrictions and flexibilities to allow a sophisticated performance practice to emerge?

For the performer he/she is performing on a new instrument which is composed of the complete system acoustic-instrument-plus- electronic-network. Any new instrument takes time to master. Hence there is a danger that a piece for electronically processed acoustic instrument will fall short of our musical expectations because no matter how good the performer, his or her mastery of the new system is unlikely to match his or her mastery of the acoustic instrument *alone* with the centuries of performance practice from which it arises. There is a danger that electronic extension may lead to musical trivialisation. Because success in this sphere depends on a marriage of good instrument design and evolving performance practice, it takes time! From this perspective it might be best to establish a number of sophisticated electronic-extension-archetypes which performers could, in time, learn to master as a repertoire for these new instruments develops.

The second paradigm is that of pure electro-acoustic (studio) composition. Such composition may be regarded as suffering from the disadvantage that it is not a 'before your very eyes', and interpreted medium. In this respect it has the disadvantages, but also the advantages, that film has viz-a-viz theatre. What film lacks in the way of reinterpretable archetypicality(!) it makes up for in the closely observed detail of location and specifically captured human uniqueness. Similarly, studio composition can deal with the uniqueness of sonic events and with the conjuring of alternative or imaginary sonic landscapes outside the theatre of musical performance itself. Moreover, sound diffusion adds an element of performance and interpretation (relating a work to the acoustic environment in which it is presented) not available in the presentation of cinema.

However, electro-acoustic composition does present us with an entirely new dilemma. Performed music works on sound archetypes. When I finger an "mf E" on a flute, I expect to hear an example of a class of possible sounds which all satisfy the restrictions placed on being a flute "mf E". Without this fact there can be no interpretation of a work. However, for a studio composer, every sound may be treated as a unique event. Its unique properties are reproducible and can be used as the basis of compositional processes which depend on those uniquenesses. The power of the computer to both record and transform *any* sound whatsoever, means that a "performance practice" (so to speak) in the traditional sense is neither necessarily attainable or desirable. But as a result the studio composer must take on many of the functions of the performer in sound production and articulation.

This does not necessarily mean, however, that all aspects of sound design need to be explicitly understood. As argued above, the success of studio produced sound-art depends on the fusion of the roles of composer and performer in the studio situation. For this to work effectively, real-time processing (wherever this is feasible) is a desirable goal.

In many studio situations in which I have worked, in order to produce some subtly time-varying modification of a sound texture or process, I had to provide to a program a file listing the values some parameter will take and the times at which it will reach those values (a *breakpoint table*). I then ran the program and subsequently heard the result. If I didn't like this result, I had to modify the values in the table and run the program again, and so on. It is clearly simpler and more efficacious, when first exploring any time-varying process and its effect on a particular sound, to move a fader, turn a knob, bow a string, blow down a tube (etc) and hear the result of this physical action as I make it. I can then adjust my physical actions to satisfy my aural experience - I can explore intuitively, without going through any conscious explanatory control process.

In this way I can learn intuitively, through performance, what is "right" or "wrong" for me in this new situation without necessarily having any conscious explicit knowledge of how I made this decision. The power of the computer both to generate and provide control over the whole sound universe does not require that I explicitly know where to travel.

But beyond this, if the program also recorded my hand movements I could store these intuitively chosen values and these could then be reapplied systematically in new situations, or analysed to determine their mathematical properties so I could consciously generalise my intuitive knowledge to different situations.

In this interface between intuitive knowledge embedded in bodily movement and direct aural feedback, and explicit knowledge carried in a numerical representation, lies one of the most significant contributions of computing technology to the process of musical composition, blurring the distinction between the skill of the performer (intuitive knowledge embedded in performance practice) and the skill of the composer and allowing us to explore the new domain in a more direct and less theory-laden manner.

SOUND COMPOSITION : AN OPEN UNIVERSE

Finally, we must declare that the realm of sound composition is a dangerous place for the traditional composer! Composers have been used to working within domains established for many years or centuries. The tempered scale became established in European music over 200 years ago, as did many of the instrumental families. New instruments (e.g. the saxophone) and new ordering procedures (e.g. serialism) have emerged and been taken on board gradually.

But these changes are minor compared with the transition into sonic composition where every sound and every imaginable process of transformation is available. The implication we wish the reader to draw is that this new domain may not merely lack established limits at this moment, it may turn out to be intrinsically unbounded. The exploratory researching approach to the medium, rather than just the mastering and extension of established craft, may be a necessary requirement to come to grips with this new domain.

Unfortunately, the enduring academic image of the European composer is that of a man (sic) who, from the depths of his (preferably Teutonic) wisdom, wills into existence a musical score out of pure thought. In a long established and stable tradition with an almost unchanging set of sound-sources (instruments) and a deeply embedded performance practice, perhaps these supermen exist. In contrast, however, good sound composition always includes a process of discovery, and hence a coming to terms with the unexpected and the unwilled, a process increasingly informed by experience as the composer engages in her or his craft, but nevertheless always open to both surprising discovery and to errors of judgement. Humility in the face of experience is an essential character trait!

In particular the sound examples in this book are the result of applying particular musical tools *to particular sound sources*. Both must be selected with care to achieve some desired musical goal. One cannot simply apply a process, "turn the handle", and expect to get a perceptually similar transformation with whatever sound source one puts into the process. Sonic Art is not like arranging.

We already know that sound composition presents us with many modes of material-variation which are unavailable - at least in a specifically controllable way - in traditional musical practice. It remains to be seen whether the musical community will be able to draw a definable boundary around available techniques and say "this is sound composition as we know it" in the same sense that we can do this with traditional instrumental composition.

CHAPTER 1

THE NATURE OF SOUND

SOUNDS ARE NOT NOTES

One of the first lessons a sound composer must learn is this. Sounds are not notes. To give an analogy with sculpture, there is a fundamental difference between the idea "black stone" and the particular stone I found this morning which is black enough to serve my purposes. This particular stone is a unique piece of material (no matter how carefully I have selected it) and I will be sculpting with this unique material. The difficulty in music is compounded by the fact that we discuss musical form in terms of idealisations. "F5 mf" on a flute is related to "E-flat4 ff" on a clarinet. But these are relations between ideals, or classes, of sounds. For every "F5 mf" on a flute is a different sound-event. It's particular grouping of micro-fluctuations of pitch, loudness and spectral properties is unique.

Traditional music is concerned with the relations among certain abstracted properties of real sounds - the pitch, the duration, the loudness. Certain other features like pitch stability, tremolo and vibrato control etc. are expected to lie within perceptible limits defined by performance practice, but beyond this enter into a less well-defined sphere known as interpretation. In this way, traditionally, we have a structure defined by relations among archetypal properties of sounds and a less well-defined aura of acceptability and excellence attached to other aspects of the sound events.

With sound recording, the unique characteristics of the sound can be reproduced. This has innumerable consequences which force us to extend, or divert from, traditional musical practice.

To give just 2 examples...

1. We can capture a very special articulation of sound that cannot be reproduced through performance - the particular tone in which a phrase is spoken on a particular occasion by an untrained speaker; the resonance of a passing wolf in a particular forest or a particular vehicle in a road-tunnel; the ultimate extended solo in an all-time great improvisation which arose out of the special combination of performers involved and the energy of the moment.

2. Exact reproducibility allows us to generate transformations that would otherwise be impossible. For example, by playing two copies of a sound together with a slight *delay* before the onset of the second we generate a pitch because the exact repetition of sonic events at an exact time interval corresponds to a fixed frequency and hence a perceived pitch.

The most important thing to understand, however, is that a sound is a sound is a sound is a sound. It is not an example of a pitch class or an instrument type. It is a unique object with its own particular properties which may be revealed, extended and transformed by the process of sound composition.

Furthermore, sounds are multi-dimensional phenomena. Almost all sounds can be described in terms of grain (particularly onset-grain), pitch or pitch-band, pitch motion, spectral harmonicity-inharmonicity and its evolution, spectral contour and formants (See Chapter 3) and their evolution, spectral stability and its evolution, and dispersive, undulating and/or forced continuation (see Chapter 4), *all at the same time*. In dividing up the various properties of sound, we don't wish to imply that there are different classes of sounds corresponding to the various chapters in the book. In fact, sounds *can* be grouped into different classes, with fuzzy boundaries, but most sounds have most of the properties that we will discuss. As compositional tools may affect two or more perceptual aspects of a sound, as we go through the book it will be necessary to refer more than once to many compositional processes as we examine their effects from different perceptual perspectives.

UNIQUENESS AND MALLEABILITY : FROM ARCHITECTURE TO CHEMISTRY

To deal with this change of orientation, our principal metaphor for musical composition must change from one of architecture to one of chemistry.

In the past, composers were provided, by the history of instrument technology, performance practice and the formalities of notational conventions (including theoretical models relating notatables like pitch and duration) with a pool of sound resources from which musical "buildings" could be constructed. Composition through traditional instruments binds together as a class large groups of sounds and evolving–shapes–of–sounds (morphologies) by collecting acceptable sound types together as an "instrument" e.g. a set of struck metal strings (piano), a set of tuned gongs (gamelan), and uniting this with a tradition of performance practice. The internal shapes (morphologies) of sound events remain mainly in the domain of performance practice and are not often subtly accessed through notation conventions. Most importantly however, apart from the field of percussion, the overwhelming dominance of pitch as a defining parameter in music focuses interest on sound classes with relatively stable spectral and frequency characteristics.

We might imagine an endless beach upon which are scattered innumerable unique pebbles. The previous task of the instrument builder was to seek out all those pebbles that were completely black to make one instrument, all those that were completely gold to make a second instrument, and so on. The composer then becomes an expert in constructing intricate buildings in which every pebble is of a definable colour. As the Twentieth Century has progressed and the possibilities of conventional instruments have been explored to their limit we have learned to recognise various shades of grey and gilt to make out architecture ever more elaborate.

Sound recording, however, opens up the possibility that any pebble on the beach might be usable – those that are black with gold streaks, those that are multi–coloured. Our classification categories are overburdened and our original task seems to become overwhelmed. We need a new perspective to understand this new world.

In sonic terms, not only sounds of indeterminate pitch (like unpitched percussion, definable portamenti, or inharmonic spectra) but those of unstable, or rapidly varying spectra (the grating gate, the human speech–stream) must be accepted into the compositional universe. Most sounds simply do not fall into the neat categories provided by a pitched–instruments oriented conception of musical architecture. As most traditional building plans used pitch (and duration) as their primary ordering principles, working these newly available materials is immediately problematic. A complete reorientation of musical thought is required – together with the power provided by computers – to enable us to encompass this new world of possibilities.

We may imagine a new personality combing the beach of sonic possibilities, not someone who selects, rejects, classifies and measures the acceptable, but a chemist who can take any pebble and by

numerical sorcery, separate its constituents, merge the constituents from two quite different pebbles and, in fact, transform black pebbles into gold pebbles, and vice versa.

The signal processing power of the computer means that *sound itself* can now be manipulated. Like the chemist, we can take apart what were once the raw materials of music, reconstitute them, or transform them into new and undreamt of musical materials. Sound becomes a fluid and entirely malleable medium, not a carefully honed collection of givens. Sculpture and chemistry, rather than language or finite mathematics, become appropriate metaphors for what a composer might do, although mathematical and physical principals will still enter strongly into the design of both musical tools and musical structures.

The precision of computer signal processing means, furthermore, that previously evanescent and uncontainable features of sounds may be analysed, understood, transferred and transformed in rigorously definable ways. A minute audible feature of a particular sound can be magnified by time–stretching or brought into focus by cyclic repetition (as in the works of Steve Reich). The evolving spectrum of a complex sonic event can be pared away until only a few of the constituent partials remain, transforming something that was perhaps coarse and jagged, into something aetherial (*spectral tracing* : see Chapter 3). We may exaggerate or contradict – in a precise manner – the energy (loudness) trajectory (or envelope) of a sound, enhancing or contradicting its gestural propensities, and we can pass between these "states of being" of the sound with complete fluidity, tracing out an audible path of musical connections – a basis for musical form– building.

This shift in emphasis is as radical as is possible – from a finite set of carefully chosen archetypal properties governed by traditional "architectural" principles, to a continuum of unique sound events and the possibility to stretch, mould and transform this continuum in any way we choose, to build new worlds of musical connectedness. To get any further in this universe, we need to understand the properties of the "sonic matter" with which we must deal.

THE REPRESENTATION OF SOUND – PHYSICAL & NUMERICAL ANALOGUES

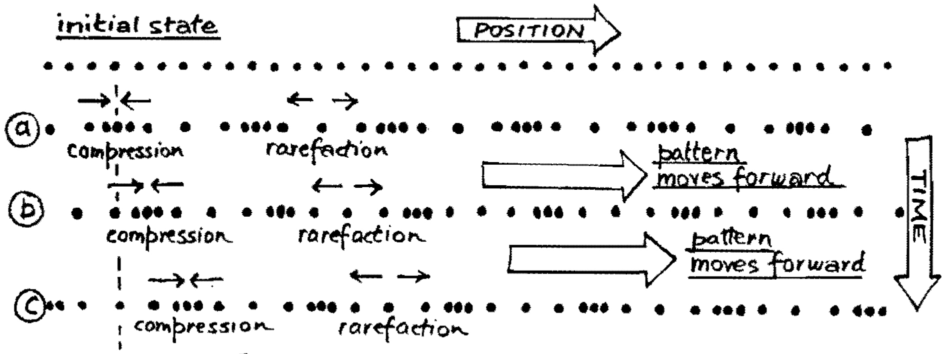
To understand how this radical shift is possible, we must understand both the nature of sound, and how it can now be physically represented. Fundamentally sound is a pressure wave travelling through the air. Just as we may observed the ripples spreading outwards from a stone thrown into still water, when we speak similar ripples travel through the air to the ears of listeners. And as with all wave motion, it is the pattern of disturbance which moves forward, rather than the water or air itself. Each pocket of air (or water) may be envisaged as vibrating about its current position and passing on that vibration to the pocket next to it. This is the fundamental difference between the motion of sound in air and the motion of the wind where the air molecules move en masse in one direction. It also explains how sound can travel very much faster than the most ferocious hurricane.

Sound waves in air differ from the ripples on the surface of a pond in another way. The ripples on a pond (or waves on the ocean) disturb the surface at right angles (up and down) to the direction of motion of that wave (forward). These are therefore known as lateral waves. In a sound wave, the air is alternatively compressed and rarefied in the same direction as the direction of motion (See **Diagram 1**).

However, we can represent the air wave as a graph of pressure against time. In such a graph we represent pressure on the vertical axis and time on the horizontal axis. So our representation ends up looking just like a lateral wave! (See **Diagram 1**).

SOUND WAVES

Sound waves are patterns of compression and rarefaction of the air, in a direction parallel to the direction of motion of the wave.



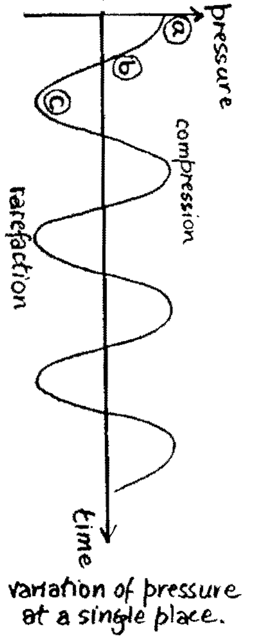


DIAGRAM 1

The representation of a sound wave as a graph of pressure against time looks like a transverse wave.

TRANSVERSE WAVES

Waves on the surface of a pond involve motion of the water in a direction perpendicular to the direction of motion of the waves.

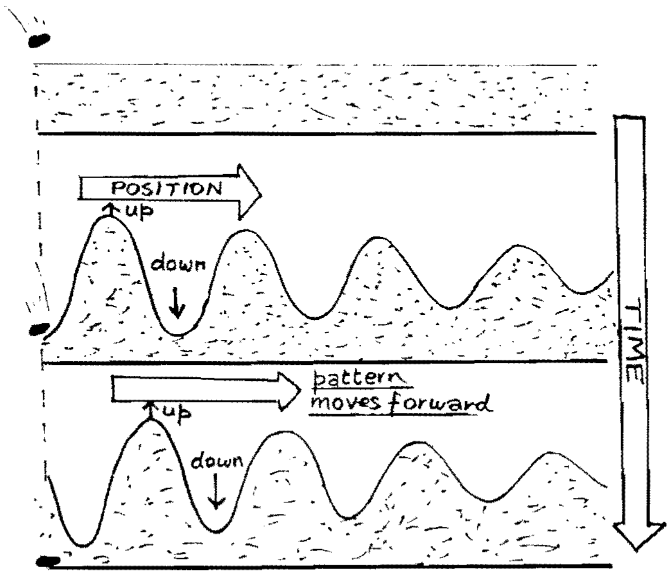
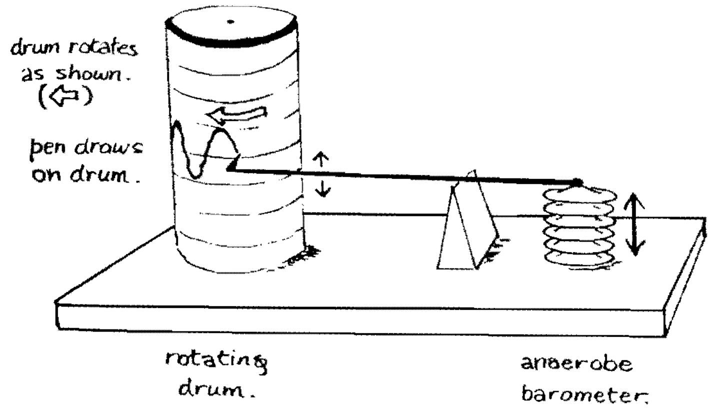


DIAGRAM 2



Anaerobe barometer stack expands and contacts as air pressure varies over time.
Point of pen traces out similar movements, which are recorded as a (spatial) trace on the regularly rotating drum.

Such waves are, of their very nature, ephemeral. Without some means to 'halt time' and to capture their form out of time, we cannot begin to manipulate them. Before the Twentieth Century this was technologically impossible. Music was reproducible through the continuity of instrument design and performance practice and the medium of the score, essentially a set of instructions for producing sounds anew on known instruments with known techniques.

The trick involved in capturing such ephemeral phenomena is the conversion of time information into spatial information. a simple device which does this is the chart-recorder, where a needle traces a graph of some time-varying quantity on a regularly rotating drum. (See **Diagram 2**).

And in fact the original phonograph recorder used exactly this principle, first converting the movement of air into the similar (analogue) movements of a tiny needle, and then using this needle to scratch a pattern on a regularly rotating drum. In this way, a pattern of pressure in time is converted to a physical shape in space. The intrinsically ephemeral had been captured in a physical medium.

This process of creating a spatial analogue of a temporal event was at the root of all sound-recording before the arrival of the computer and is still an essential part of the chain in the capture of the phenomenon of sound. The fundamental idea here is of an analogue. When we work on sound we in fact work on a spatial or electrical or numerical analogue, a very precise copy of the form of the sound-waves in an altogether different medium.

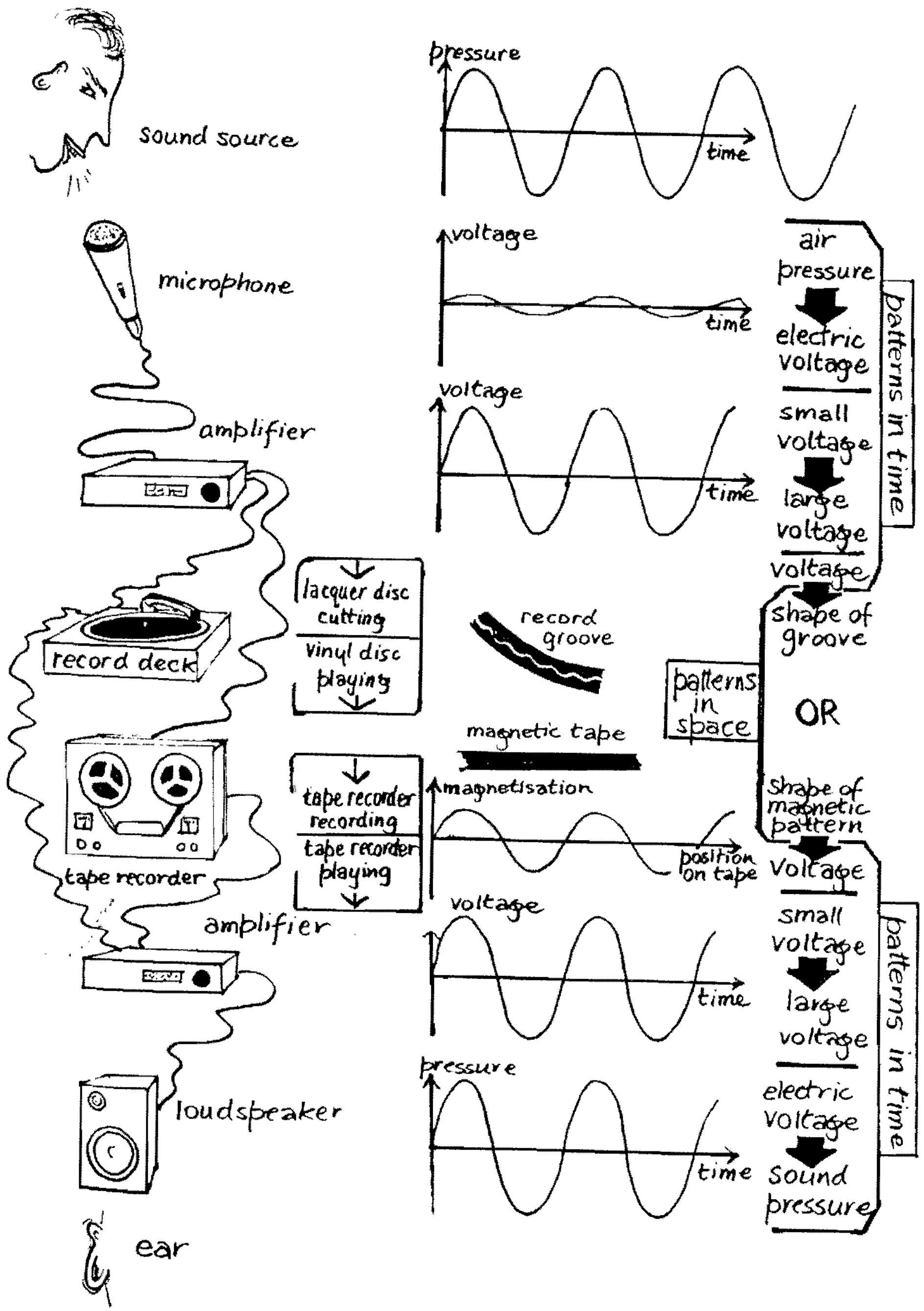
The arrival of electrical power contributed greatly to the evolution of an Art of Sound itself. First of all, reliable electric motors ensured that the time-varying wave could be captured and reproduced reliably, as the rotating devices used could be guaranteed to run at a constant speed. More importantly, electricity itself proved to be the ideal medium in which to create an analogue of the air pressure-waves. Sound waves may be converted into electrical waves by a microphone or other transducer. Electrical waves are variations of electrical voltage with time - analogues of sound waves. Such electrical analogues may also be created using electrical oscillators, in an (analogue) synthesizer.

Such electrical patterns are, however, equally ephemeral. They must still be converted into a spatial analogue in some physical medium if we are to hold and manipulate sound-phenomena out of time. In the past this might be the shape of a groove in rigid plastic (vinyl discs) or, the variation of magnetism along a tape (analogue tape recorder). By running the physical substance past a pickup (the needle of a record player, the head of a tape recorder) at a regular speed, the position information was reconverted to temporal information and the electrical wave was recreated. The final link in the chain is a device to convert electrical waves back into sound waves. This is the loudspeaker. (See **Diagram 3**).

Note that in all these cases we are attempting to preserve the shape, the form, of the original pressure wave, though in an entirely different medium.

The digital representation of sound takes one further step which gives us ultimate and precise control over the very substance of these ephemeral phenomena. Using a device known as an A to D (analogue to digital) converter, the pattern of electrical fluctuation is converted into a pattern of numbers. (See **Appendix p1**). The individual numbers are known as samples (not to be confused with chunks of recorded sounds, recorded in digital samplers, also often referred to as 'samples'), and each one represents the instantaneous (almost!) value of the pressure in the original air pressure wave.

DIAGRAM 3



We now arrive at an essentially abstract representation of sonic substance for, although these numbers are normally stored in the spatial medium of magnetic domains on a computer disk, or pits burned in a CD, there need be no simple relationship between their original temporal order and the physical–spatial order in which they are stored. Typically, the contents of a file on a hard disk are scattered over the disk according to the availability of storage space. What the computer can do however is to re–present to us, in their original order, the numbers which represent the sound.

Hence what was once essentially physical and temporally ephemeral has become abstract. We can even represent the sound by writing a list of sample values (numbers) on paper, providing we specify the sample rate (how many samples occur at a regular rate, in each second), though this would take an inordinately long time to achieve in practice.

It is interesting to compare this new musical abstraction with the abstraction involved in traditional music notational practice. Traditional notation was an abstraction of general and easily quantifiable large–scale properties of sound events (or performance gestures, like some ornaments), represented in written scores. This abstracting process involves enormous compromises in that it can only deal accurately with finite sets of definable phenomena and depends on existing musical assumptions about performance practice and instrument technology to supply the missing information (i.e. most of it!) (See the discussion in *On Sonic Art*). In contrast, the numerical abstraction involved in digital recording leaves nothing to the imagination or foreknowledge of the musician, but consequently conveys no abstracted information on the macro level.

THE REPRESENTATION OF SOUND _ THE DUALITY OF TIME AND FREQUENCY

So far our discussion has focused on the representation of the actual wave–movement of the air by physical and digital analogues. However, there is an alternative way to think about and to represent a sound. Let us assume to begin with that we have a sound which is not changing in quality through time. If we look at the pressure wave of this sound it will be found to repeat its pattern regularly (See **Appendix** p3).

However, perceptually we are more interested in the perceived properties of this sound. Is it pitched or noisy (or both)? Can we perceive pitches within it? etc. In fact the feature underlying these perceived properties of a static sound is the disposition of its partials. These may be conceived of as the simpler vibrations from which the actual complex waveform is constructed. It was Fourier (an Eighteenth Century French mathematician investigating heat distribution in solid objects) who realised that any particular wave–shape (or, mathematically, any function) can be recreated by summing together elementary sinusoidal waves of the correct frequency and loudness (**Appendix p2**). In acoustics these are known as the partials of the sound. (**Appendix p3**).

If we have a regular repeating wave–pattern we can therefore also represent it by plotting the frequency and loudness of its partials, a plot known as the spectrum of the sound. We can often deduce perceptual information about the sound from this data. The pattern of partials will determine if the sound will be perceived as having a single pitch or several pitches (like a bell) or even (with singly pitched sounds) what its pitch might be.

The spectral pattern which produces a singly pitched sound has (in most cases) a particularly simple structure, and is said to be "harmonic". Thus should not to be confused with the traditional notion of

HArmony in European music. Thus we use double capitalisation for the latter, and none for the former. (For a fuller discussion, see Chapter 2). The bell spectrum, in contrast, is known as an inharmonic spectrum.

It is important to note that this new representation of sound is out–of–time. We have converted the temporal information in our original representation to frequency information in our new representation.

These two representations of sound are known as the time–domain representation and the frequency–domain representation respectively. The mathematical technique which allows us to convert from one to the other is known as the *Fourier transform* (and to convert back again, the *inverse Fourier transform*) which is often implemented on computers in a highly efficient algorithm called the Fast Fourier Transform (FFT).

If we now wish to represent the spectrum of a sound which varies in time (i.e. any sound of musical interest and certainly any naturally occurring sound) we must divide the sound into tiny time–snapshots (like the frames of a film) known as windows. A sequence of these windows will show us how the spectrum evolves with time (the *phase vocoder* does just this, see **Appendix p11**).

Note also that a very tiny fragment of the time–domain representation (a fragment shorter than a wavecycle), although it gives us accurate information about the time–variation of the pressure wave, gives us no information about frequency. Converting it to frequency data with the FFT will produce energy all over the spectrum. Listening to it we will hear only a click (whatever the source). Conversely, the frequency domain representation gives us more precise information about the frequency of the partials in a sound the larger the time window used to calculate it. But in enlarging the window, we track less accurately how the sound changes in time. This trade–off between temporal information and frequency information is identical to the quantum mechanical principle of indeterminacy, where the time and energy (or position and momentum) of a particle/wave cannot both be known with accuracy – we trade off our knowledge of one against our knowledge of the other.

TIME FRAMES : SAMPLES, WAVE-CYCLES, GRAINS AND CONTINUATIONS.

In working with sound materials, we quickly become aware of the different time–frames involved and their perceptual consequences. Stockhausen, in the article *How Time Passes* (published in *Die Reihe*) argued for the unity of formal, rhythmic and acoustic time–frames as the rationale for his composition *Gruppen*. This was a fruitful stance to adopt as far as *Gruppen* was concerned and fits well with the "unity" mysticism which pervades much musical score analysis and commentary, but it does not tally with aural experience.

Extremely short time frames of the order of 0.0001 seconds have no perceptual significance at all. Each sample in the digital representation of a waveform corresponds to a time less than 0.0001 seconds. Although every digitally recorded sound is made out of nothing but samples, the individual sample can tell us nothing about the sound of which it is a part. Each sample, if heard individually, would be a broad–band click of a certain loudness. Samples are akin to the quarks of subatomic particle theory, essential to the existence and structure of matter, but not separable from the particles (protons and neutrons) they constitute.

The first significant object from a musical point of view is a shape made out of samples, and in particular a wavecycle (a single wavelength of a sound). These may be regarded as the atomic units of sound. The shape and duration of the wavecycle will help to determine the properties (the spectrum and pitch) of the sound of which it is a part. But a single wavecycle is not sufficient on its own to determine these properties. As pitch depends on frequency, the number of times per second a waveform is repeated, a single wavecycle supplies no frequency information. Not until we have about six wavecycles do we begin to associate a specific pitch with the sound. Hence there is a crucial perceptual boundary below which sounds appear as more or less undifferentiated clicks, regardless of their internal form, and above which we begin to assign specific properties (frequency, pitch/noise, spectrum etc) to the sound (for a fuller discussion of this point see *On Sonic Art*).

This perceptual boundary is nicely illustrated by the process of *waveset time-stretching*. This process lengthens a sound by repeating each waveset (A waveset is akin to a wavecycle, but not exactly the same thing, For more details see **Appendix p55**). With noisy sources, the wavesets vary widely from one to another but we hear only the net result, a sound of indefinite pitch. Repeating each waveset lengthens the sound and introduces some artefacts. If we repeat each waveset five or six times however, each one is present long enough to establish a specific pitch and spectral quality and the original source begins to transform into a rapid stream of pitched beads. (**Sound example 1.0**).

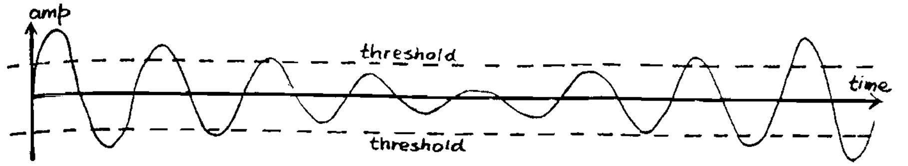
Once we can perceive distinctive qualitative characteristics in a sound, we have a *grain*. The boundary between the wavecycle time frame and the grain time-frame is of great importance in instrument design. For example, imagine we wished to separate the grains in a sound (like a rolled "R") by examining the loudness trajectory of the sound. Intuitively we can say that the grains are the loud part of the signal, and the points between grains the quietest parts. If we set up an instrument which waits for the signal level to drop below a certain value (a threshold) and then cuts out the sound (*gating*), we should be able to separate the individual grains.

However, on reflection, we see that this procedure will not work. The instantaneous level of a sound signal constantly varies from positive to negative so, at least twice in every wavecycle, it will fall below the threshold and our procedure will chop the signal into its constituent half wavecycles or smaller units (see **Diagram 4**) – not what we intended. What we must ask the instrument to do is search for a point in the signal where the signal stays below the (absolute) gate value for a significant length of time. This time is at least of grain time-frame proportions. (See **Diagram 5**).

A Grain differs from any larger structure in that we cannot perceive any resolvable internal structure. The sound presents itself to us as an indivisible unit with definite qualities such as pitch, spectral contour, onset characteristics (hard-edged, soft-edged), pitchy/noisy/gritty quality etc. Often the grain is characterised by a unique cluster of properties which we would be hard pressed to classify individually but which enables us to group it in a particular type e.g. unvoiced "k", "t", "p", "d".

Similarly, the spectral and pitch characteristics may not be easy to pin down, e.g. certain drums have a focused spectrum which we would expect from a pitched sound (they definitely don't have noise spectra as in a hi-hat), yet no particular pitch may be discernible. Analysis of such sounds may reveal either a very short inharmonic spectrum (which has insufficient time to register as several pitches, as we might hear out in an inharmonic bell sound), or a rapidly portamentoing pitched spectrum. Although the internal structure of such sounds is the cause of what we hear, we do not resolve this internal structure in our perception. The experience of a grain is indivisible.

DIAGRAM 4



Set the sound to zero
wherever its absolute value falls below
the level of the gate

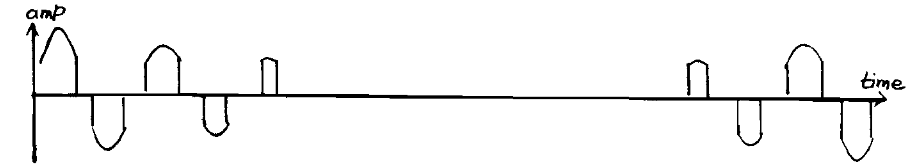
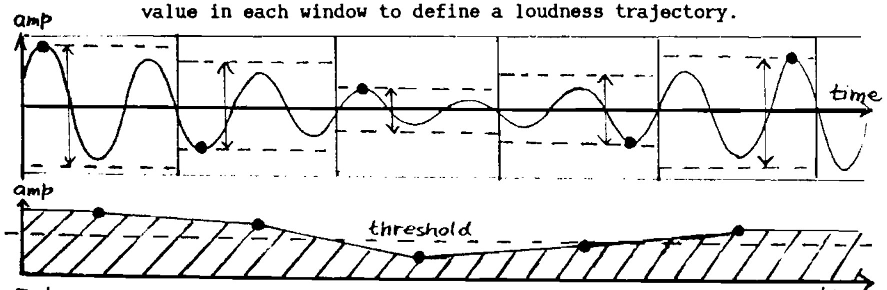
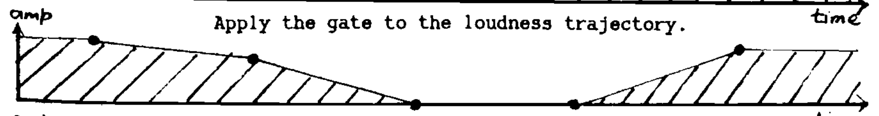


DIAGRAM 5

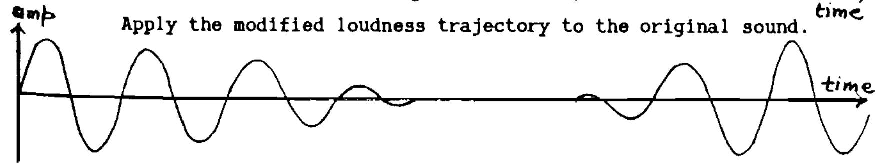
Divide up the sound into finite length windows, and use the absolute maximum value in each window to define a loudness trajectory.



Apply the gate to the loudness trajectory.



Apply the modified loudness trajectory to the original sound.



The internal structure of grains and their indivisibility was brought home to me through working with birdsong. A particular song consisted of a rapid repeated sound having a "bubbly" quality. One might presume from this that the individual sounds were internally portamentoed at a rate too fast to be heard. In fact, when slowed down to 1/8th speed, each sound was found to be comprised of a rising scale passage followed by a brief portamento!

Longer sound events can often be described in terms of an onset or attack event and a continuation. The onset usually has the timescale and hence the indivisibility and qualitative unity of a grain and we will return to this later. But if the sound persists beyond a new time limit (around .05 seconds) we have enough information to detect its temporal evolution, we become aware of movements of pitch or loudness, or evolution of the spectrum. The sound is no longer an indivisible grain: we have reached the sphere of Continuation.

This is the next important time-frame after Grain. It has great significance in the processing of sounds. For example, in the technique known as *brassage*, we chop up a sound into tiny segments and then *splice* these back together again. If we retain the order of the segments using overlapping segments from the original sound, but don't overlap them (so much) in the resulting sound, we will clearly end up with a longer sound (**Appendix p44**).

If we try to make segments smaller than the grain-size, we will destroy the signal because the *splices* (cross-fading) between each segment will be so short as to break up the continuity of the source and destroy the signal characteristics. For example, attempting to time-stretch by a factor of 2 we will in fact resplice together parts of the waveform itself, to make a waveform twice as long, and our sound will drop by an octave, as in *tape-speed variation*. If the segments are in the grain time-frame, the instantaneous pitch will be preserved, as the waveform itself will be preserved intact, and we should achieve a time-stretching of the sound without changing its pitch (the *harmoniser* algorithm). If the segments are longer than grains, their internal structure will be heard out and we will begin to notice echo effects as the perceived continuations are heard to be repeated. Eventually, we will produce a collage of motifs or phrases cut from the original material, as the segment size becomes very much larger than the grain time-frame. (**Sound example 1.1**).

The grain/continuation time-frame boundary is also of crucial importance when a sound is being time-stretched and this will be discussed more fully in Chapter 11.

The boundaries between these time-frames (wavecycle, grain, continuation) are not, of course, completely clear cut and interesting perceptual ambiguities occur if we alter the parameters of a process so that it crosses these time-frame thresholds. In the simplest case, gradually time-stretching a grain gradually makes its internal structure apparent (See Chapter 11) so we can pass from an indivisible qualitative event to an event with a clearly evolving structure of its own. Conversely, a sound with clear internal structure can be time-contracted into a structurally irresolvable grain. (**Sound example 1.2**).

Once we reach the sphere of continuation, perceptual descriptions become more involved and perceptual boundaries, as our time frame enlarges, less clear cut. If the spectrum has continuity, perception of continuation may be concerned with the morphology (changing shape) of the spectrum, with the articulation of pitch (vibrato, jitter), loudness (tremolo), spectral contour or formant gliding (see Chapter 3) etc, etc. The continuation may, however, be discontinuous as in iterative sounds (grain-streams – such as rolled "R", and low contrabassoon notes which are perceived partly as a rapid sequence of onsets) or segmented sounds (see below), or granular texture streams (e.g. maracas

with coarse pellets) where the sound is clearly made up of many individual randomised attacks. Here, variation in grain or segment speed or density and grain or segment qualities will also contribute to our aural experience of continuation.

As our time-frame lengthens, we reach the sphere of the Phrase. Just as in traditional musical practice, the boundary between a long articulation and a short phrase is not easy to draw. This is because we are no longer dealing with clear cut *perceptual* boundaries, but questions of the interpretation of our experience. A trill, without variation, lasting over four bars may be regarded as a note-articulation (an example of continuation) and may exceed in length a melodic phrase. But a trill with a marked series of loudness and speed changes might well function as a musical phrase (depending on context). (**Sound example 1.3**).

A similar ambiguity applies in the sound domain with an added difficulty. Whereas it will usually be quite clear what is a note event and what is a phrase (ornaments and trills blurring this distinction), a sound event can be arbitrarily complex. We might, for example, start with a spoken sentence, a phrase-time-frame object, then time-shrink it to become a sound of segmented morphology (See Chapter 11). As in traditional musical practice, the recognition of a phrase as such, will depend on musical context and the fluidity of our new medium will allow us to shrink, or expand, from one time-frame to another.

A similar ambiguity applies as we pass further up the time-frame ladder towards larger scale musical entities. We can however *construct* a series of nested time-frames up to the level of the duration of an entire work. These nested time-frames are the basis of our perception of both rhythm and larger scale form and this is more fully discussed in Chapter 9.

THE SOUND AS A WHOLE – PHYSICALITY AND CAUSALITY

Most sounds longer than a grain can be described in terms of an onset and a continuation. A detailed discussion of the typology of sounds can be found in *On Sonic Art* and in the writings of the Groupe de Recherches Musicales. Here, I would like to draw attention to two aspects of our aural experience.

The way in which a sound is attacked and continues provides evidence of the physicality of its origin. In the case of transformed or synthesized sounds, this evidence will be misleading in actuality, but we still gain an impression of an *imagined* origin of the sound.

It is important to bear this distinction in mind. As Pierre Schaeffer was at pains to stress, once we begin working with sounds as our medium the *actual* origin of those sounds is no longer of any concern. This is particularly true in the era of computer sound transformation. However, the *apparent* origin (or physicality) of the sound remains an important factor in our perception of the sound in whatever way it is derived.

We may look at this in another way. With the power of the computer, we can transform sounds in such radical ways that we can no longer assert that the goal sound is related to the source sound merely because we have derived one from the other. We have to establish a connection in the experience of the listener either through clear spectral, morphological, or etc similarities between the sounds, or a clear path through a series of connecting sounds which gradually change their characteristics from those of the source, to those of the goal. This is particularly true when the *apparent* origin (physicality) of the goal sound is quite different to that of the source sound.

Thus, for example, we may *spectrally time-stretch* and change the loudness trajectory (*enveloping*) of a vocal sound, producing wood-like attacks, which are then progressively distorted to sound like unpitched drum sounds. **(Sound example 1.4).**

In general, sounds may be activated in two ways - by a single physical event (e.g. a striking blow), or by a continuous physical event (blowing, bowing, scraping). In the first case, the sound may be internally damped producing a short sound, or grain – a xylophone note, a drum-stroke, a vocal click. It may be short, but permitted to resonate through an associated physical system, e.g. a cello soundbox for a pizzicato note, a resonant hall for a drum stroke. Or the material itself may have internal resonating properties (bells, gongs, metal tubes) producing a gradually attenuated continuum of sound.

In the case of continuous excitation of a medium, the medium may resonate, producing a steady pitch which varies in loudness with the energy of the excitation e.g. Flute, Tuba, Violin. The medium may vibrate at a frequency related to the excitation force (e.g. a rotor-driven siren, or the human voice in some circumstances) so that a varying excitation force varies the pitch. Or the contact between exciting force and vibrating medium may be discontinuous, producing an iterated sound (rolled "R", drum roll etc).

The vibrating medium itself may be elastically mobile – a flexatone, a flexed metal sheet, a mammalian larynx – so that the pitch or spectrum of the sound varies through time. The material may be only gently coaxed into motion (the air in a "Bloogle", the shells in a Rainmaker) giving the sound a soft onset, or the material may be loosely bound and granular (the sand or beads in a Shaker or Wind-machine) giving the sound a diffuse continuation. Resonating systems will stabilise after a variety of transient or unstable initiating events (flute breathiness, coarse hammer blows to a metal sheet) so that a complex and disconnected onset leads to a stable or stably evolving spectrum.

I am not suggesting that we consciously analyse our aural experience in this way. On the contrary, aural experience is so important to us that we already have an intuitive knowledge (see earlier) of the physicality of sound-sources. I also do not mean that we see pictures of physical objects when we hear sounds, only that our aural experience is grounded in physical experience in a way which is not necessarily consciously understood or articulated. Transforming the characteristics of a sound-source automatically involves transforming its perceived physicality and this may be an important feature to bear in mind in sound composition.

In a similar and not easily disentangled way, the onset (or attack) properties of a sound give us some inkling of the cause of that sound – a physical blow, a scraping contact, a movement, a vocal utterance. The onset or attack of a sound is always of great significance if only because it is the moment of greatest surprise when we know nothing about the sound that is to evolve, whereas during the continuation phase of the sound, we are articulating what the onset has revealed. It is possible to give the most unlikely sounds an apparent vocal provenance by very carefully *splicing* a vocal onset onto a non-vocal continuation. The vocal "causality" in the onset can adhere to the ensuing sound in the most unlikely cases.

STABILITY & MOTION : REFERENCE FRAMES

In general, any property of a sound (pitch, loudness, spectral contour etc) may be (relatively) stable, or it may be in motion (portamento, crescendo, opening of a *filter*, etc.) Furthermore, motion may itself be in motion, e.g. cyclically varying pitch (vibrato) may be accelerating in cycle-duration, while shrinking in pitch-depth. Cyclical motions of various kinds (tremolo, vibrato, spectral vibrato, etc.) are often regarded as stable properties of a sound.

We may be concerned with stable properties, or with the nature of motion itself and these aspects of sound properties are usually perceptually distinguishable. Their separation is exaggerated in Western Art Music by the use of a *discontinuous* notation system which notates static properties well, but moving properties much less precisely (for a fuller discussion see *On Sonic Art*).

Furthermore, our experiences of sonic change are often fleeting and only inexactly reproducible. We can, for example, reproduce the direction, duration and range of a rising portamento and in practice, we can differentiate a start-weighted from an end-weighted portamento. (See **Diagram 6**).

In many non-Western cultures, subtle control of such distinctions (portamento type, vibrato speed and depth etc) are required skills of the performance practice. But the reproduction of a complex portamento trajectory (see **Diagram 7**) with any precision would be difficult merely from immediate memory. The difficulty of immediate reproducibility makes repetition in performance very difficult and therefore acquaintance and knowledge through familiarity impossible to acquire.

With computer technology, however, complex, time-varying aspects of a sound-event can be tied down with precision, reproduced and transferred to other sounds. For the first time, we have sophisticated control over sonic motion in many dimensions (pitch, loudness, spectral contour or formant evolution, spectral harmonicity/inharmonicity etc. etc.) and can begin to develop a discipline of motion itself (see Chapter 13).

In sound composition, the entire continuum of sound possibilities is open to us, and types of motion are as accessible as static states. But in our perception of our sound universe there is another factor to be considered, that of the *reference-frame*. We may choose to (it is not inevitable!) organise a sound property in relation to a set of reference values. These provide reference points, or nameable foci, in the continuum of sonic possibilities. Thus, for example, any particular language provides a phonetic reference-frame distinguishing those vowels and consonant types to be regarded as different (and hence capable of articulating different meanings) within the continuum of possibilities. These distinctions may be subtle ("D" and "T" in English) and are often problematic for the non-native speaker (English "L" and "R" for a Japanese speaker).

Usually, these reference frames refer to stable properties of the sonic space but this is not universal. Consonants like "W" and "Y" (in English) and various vowel diphthongs are in fact defined by the motion of their spectral contours (formants : see **Appendix p3**). But in general, reference frames for motion types are not so well understood and certainly do not feature strongly in traditional Western art music practice. Nevertheless, we are equipped in the sphere of language perception, at the level of the phoneme ("w", "y", etc. and tone languages like Chinese) and of "tone-of-voice", to make very subtle distinctions of pitch motion and spectral motion. They are vital to our comprehension at the phonemic and semantic level. And in fact, sounds with moving spectral contours tend to be classified alongside sounds with stable spectral contours in the phonetic classification system of any particular language.

DIAGRAM 6

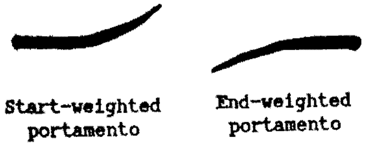


DIAGRAM 7

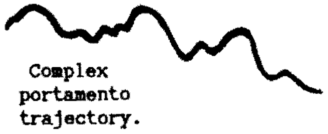
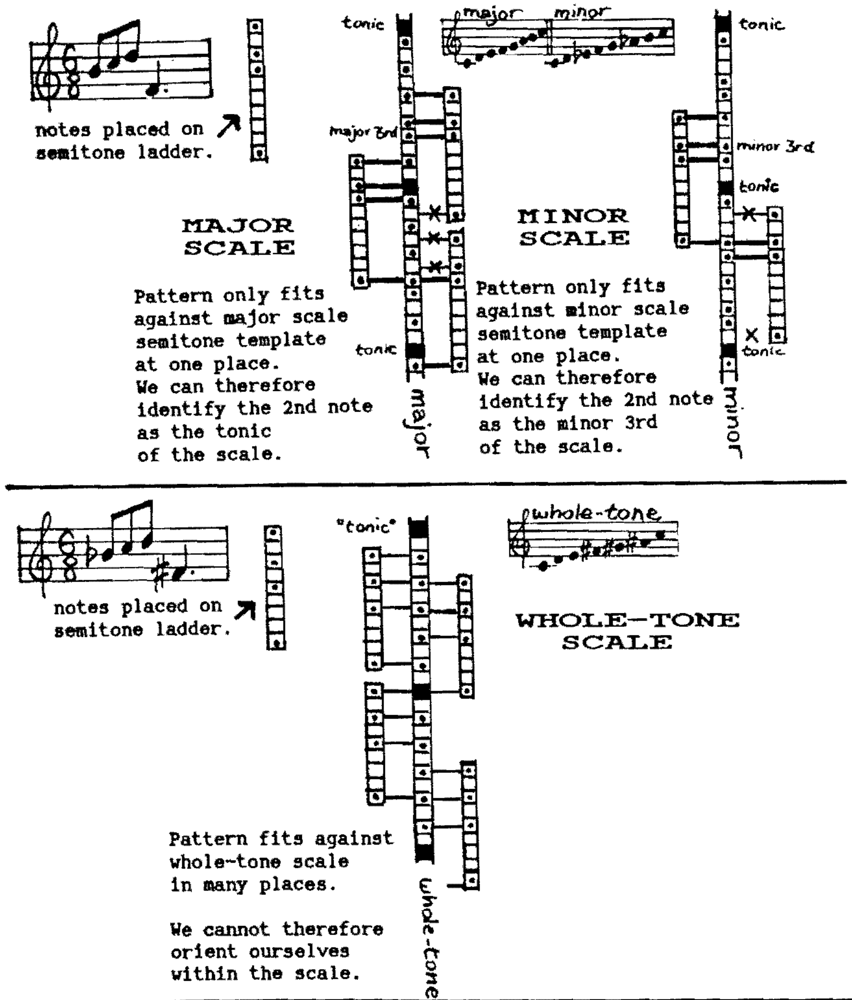


DIAGRAM 8



We can see the same mixed classification in the Indian raga system where a raga is often defined in terms of both a scale (of static pitch values) and various motivic figures often involving sliding intonations (motion–types). In general, in the case of pitch reference–frames, values lying off the reference frame may only be used as ornamental or articulatory features of the musical stream. They have a different status to that of pitches on the reference frame (cf. "blue notes" in Jazz, certain kinds of baroque ornamentation etc).

A reference frame gives a structure to a sonic space enabling us to say "I am here" and not somewhere else. But pitch reference frames have special properties. Because we normally accept the idea of octave equivalence (doubling the frequency of a pitched sound produces the "same" pitch, one octave higher), pitch reference frames are cyclic, repeating themselves at each octave over the audible range. In contrast, in the vowel space, we have only one "octave" of possibilities but we are still able to recognise, without reference to other vowel sounds, where we are in the vowel space. We have a sense of absolute position.

People with perfect pitch can deal with pitch space in this absolute sense, but for most of us, we have only some general conception of high and low. We can, however, determine our position *relative* to a given pitch, using the notion of interval. If the octave is divided by a mode or scale into an asymmetrical set of intervals, we can tell where we are from a small set of notes without hearing the key note because the interval relationships between the notes orient us within the set–of– intervals making up the scale. We cannot do this trick, however, with a completely symmetrical scale (whole–tone scale, chromatic scale) without some additional clues (See **Diagram 8**).

Cyclic repetition over the domain of reference and the notion of interval are specific to pitch and time reference–frames. However, time reference frames which enter into our perception of rhythm are particularly contentious in musical circles and I will postpone further discussion of these until Chapter 9.

Traditional Western music practice is strongly wedded to pitch reference frames. In fact on many instruments (keyboards, fretted strings) they are difficult to escape. However, in sonic composition we can work....

(a) with static properties on a reference frame.

(b) with static properties *without* a reference frame.

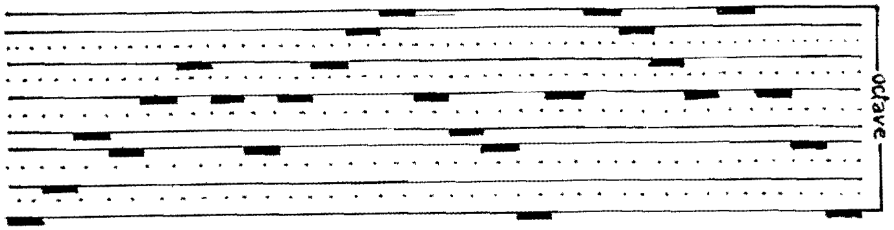
(c) with properties of motion.

Furthermore, we can transform the reference frame itself, through time, or move on to, and away from, such a frame. Computer control permits the very precise exploration of this area of new possibilities.

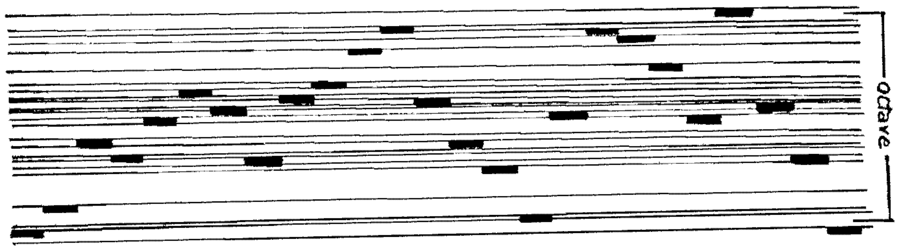
It is particularly important to understand that we can have pitch, and even stable pitch, without having a stable reference frame and hence without having a *HArmonic sense of pitch* in the traditional sense. (See **Diagram 9**). We can even imagine establishing a reference frame for pitch–motion types without having a HArmonic frame – this we already find in tone languages in China and Africa.

DIAGRAM 9

Stable pitches defining a reference frame.



Stable pitches not defining a reference frame.
(every pitch is different)



THE MANIPULATION OF SOUND : TIME-VARYING PROCESSING

Just as I have been at pains to stress that sound events have several time-varying properties, it is important to understand that the compositional processes we apply to sounds may also vary in time in subtle and complex ways. In general, any parameter we give to a process (pitch, speed, loudness, filter centre, density etc.) should be replaceable by a time varying quantity and this quantity should be continuously variable over the shortest time-scales.

In a tradition dominated by the notation of fixed values (pitch, loudness level, duration) etc, it is easy to imagine that processes themselves must have fixed values. In fact, of course, the apparently fixed values we dictate in a notation system are turned into subtly varying events by musical performers. Where this is not the case (e.g. quantised sequencer music on elementary synthesis modules) we quickly tire of the musical sterility.

The same fixed-value conception was transferred into acoustic modelling and dominated early synthesis experiments. And for this reason, early synthesized sounds governed by fixed values (or simply-changing values) suffered from a lack of musical vitality. In a similar fashion an "effects unit" used as an all-or-nothing process on a voice or instrument, is simply that, an "effect" which may occasionally be appropriate (to create a particular general acoustic ambience, for example) but will not sustain our interest as a compositional feature until we begin to articulate its temporal evolution.

A time-varying process can, on the other hand, effect a radical transformation *within* a sound, e.g. a gradually increasing (but subtly varying) vibrato, applied to a fairly static sound, gradually giving it a vocal quality; a gradual *spectral stretching* of a long sound causing its pitch to split into an inharmonic stream; and so on. Transforming a sound does not necessarily mean changing it in an all-or-nothing way. Dynamic *spectral interpolation* in particular (see Chapter 12) is a process which depends on the subtle use of time-varying parameters.

THE MANIPULATION OF SOUND : RELATIONSHIP TO THE SOURCE

The infinite malleability of sound materials raises another significant musical issue, discussed by Alain Savouret at the International Computer Music Conference at IRCAM in 1985. As we can do anything to a signal, we must decide *by listening* whether the source sound and the goal sound of a compositional process are in fact at all perceptually related, or at least whether we can define a *perceptible* route from one to the other through a sequence of intermediate sounds. In score-based music there is a tradition to claiming that a transformation which can be explained and who's results can be *seen* in a score, by definition defines a musical relationship. This mediaeval textual idealism is out of the question in the new musical domain.

An instrument which replaces every half-waveset by a single pulse equal in amplitude to the amplitude of the half-waveset, is a perfectly well-defined transformation, but reduces most sounds to a uniform crackling noise. Nevertheless, for complex real-world sounds, each crackling signal is unique at the sample level and uniquely related to its source sound. Hence no musical relationship has been established between the source sound and the transformed sound.

In sound composition a relationship between two sounds is established only through aurally perceptible similarity or relatedness, regardless of the methodological rigour of the process which transforms one sound into another.

Another important aspect of Savouret's talk was to distinguish between *source-focused transformation* (where the nature of the resulting sound is strongly related to the input sound, (e.g. *time-stretching* of a signal with a stable spectrum, retaining the onset unstretched) and *process-focused transformations* (where the nature of the resulting sound is more strongly determined by the transformation process itself, e.g. using very short time digital *delay* of a signal, superimposed on the non-delayed signal to produce delay-time related pitched ringing). There is, of course, an interesting area of ambiguity between the two extremes.

In general, process-focused transformations need to be used sparingly. Often when a new compositional technique emerges, e.g. pitch parallelism via the *harmonizer*, there is an initial rush of excitement to explore the new sound possibilities. But such process-focused transformations can rapidly become cliches.

Transformations focused in the source, however, retain the same infinite potential that the infinity of natural sound sources offer us. Sound-processing procedures which are sensitive to the evolving properties (pitch, loudness, spectral form and contour etc.) of the source-sound are those most likely to bring rich musical rewards.

CHAPTER 2

PITCH

WHAT IS PITCH ?

Certain sounds appear to possess a clear quality we call pitch. Whereas a cymbal or a snare-drum has no such property and a bell may appear to contain several pitches, a flute, violin or trumpet when played in the conventional way produces sounds that have the quality "pitch".

Pitch arises when the partials in the spectrum of the sound are in a particularly simple relation to one another, i.e. they are all whole number multiples of some fixed value, known as the fundamental. In mathematical terms, the fundamental frequency is the highest common factor (HCF) of these partials. When a spectrum has this structure it is said to be harmonic, and the individual partials are known as harmonics. But this must *not* be confused with the notion of "harmony" in traditional music. What happens when the partials are not in this simple relationship is discussed in the next Chapter. Thus the numbers 200, 300, 400, 500 are all whole number multiples of 100, which is their fundamental. The frequency of this fundamental determines the "height" or "value" of the pitch we hear.

In most cases, the fundamental frequency of such a sound is present in the sound as the frequency of the lowest partial, but this is not necessarily true (e.g. the lowest notes of the piano do not contain any partial whose frequency corresponds to the perceived pitch). It is important to understand, therefore, that the perceived pitch is a mental construct from a harmonic spectrum, and not simply a matter of directly perceiving a fundamental frequency in a spectrum. Such a frequency may not be physically present.

The most important feature of pitch perception is that the spectrum appears to fuse into a unitary percept, that of pitch, with a certain spectral quality or "timbre". This fusion is best illustrated by undoing it.

For example if we play a (synthesized) voice sound by placing the odd harmonics on the left loudspeaker and the even harmonics on the right loudspeaker we will hear one vocal sound between the two loudspeakers. This happens because of a phenomenon known as aural streaming. When sounds from two different sources enter our ears simultaneously we need some mechanism to disentangle the partials belonging to one sound from those belonging to the other. One way in which the ear is able to process the data relies on the micro-instabilities (jitter) in pitch, or tessitura, and loudness which all naturally occurring sounds exhibit. The partials derived from one sound will all jitter in parallel with one another, while those from the other sound will jitter differently but also in parallel with one another. This provides a strong clue for our brain to assign any particular partial to one or other of the source sounds.

In our loudspeaker experiment, however, we have removed this clue by maintaining synchronicity of microfluctuations between the partials coming from the two loudspeakers. Hence the ear does not unscramble the data into two separate sources. The voice remains a single percept. (**Sound example 2.1**).

If, now, we gradually add a *different* vibrato to each set of partials, the sound image will split. The ear is now able to group the 2 sets of data into two aural streams and assign two different source sounds to what it hears. The odd harmonics, say 300, 500, 700, 900, will continue to imply a fundamental of 100 cycles but will take on a clarinet type quality (clarinets produce only the odd harmonics in the spectrum) and move into one of the loudspeakers. The remaining harmonics, 200, 400, 600, 800, will be interpreted as having a fundamental at 200 (as 200 is the HCF of 200, 400, 600 and 800) and hence a second "voice" an octave higher, will appear to emanate from the other loudspeaker. Hence, with no change of spectral content, we have generated 2 pitch percepts from a single pitch percept. **(Sound example 2.2)**.

SPECTRAL AND HARMONIC CONCEPTIONS OF PITCH

Our definition of pitch leads us into conflict both with traditional conceptions and traditional terminology. First of all, to say that a spectrum is harmonic, is to say that the partials are *exact* multiples of the fundamental and this is the source of the perceptual *fusion*. Once this exact relationship is disturbed, this spectral fusion is disturbed (see next Chapter).

There are not different kinds of harmony in the spectrum. Most of the relationships we deal with *between* pitches in traditional Western "harmony" are between frequencies that do not stand in this simple relationship to one another (because our scale is tempered). They are approximations to whole number ratios which "work" in the functional context of Western harmonic musical language. But, more importantly, they are relationships between the averaged properties of sounds. An "A" and a "C" played on two flutes (or on one piano) are two distinct sound events, each having its own integrated spectrum. Each spectrum is integrated with itself because its internal microfluctuations run in parallel over all its own partials. But these microfluctuations are different to those in the other spectrum. Within a *single spectrum*, however, partials might roughly correspond to an A and a C (but in exact numerical proportions, unlike in the tempered scale) but will also have exactly parallel microfluctuations and hence fuse in our unitary perception of a much lower fundamental (e.g. an F 2 octaves below). We can draw analogies between these two domains (as some composers have done) but they are perceptually quite distinct.

To avoid confusion, we will try to reserve the words "HArmony" and "HArmonic" (capitalised as shown) to apply to the traditional concern with relations amongst notes in Western art music and we will refer to a spectrum having pitch as a pitch-spectrum or as having harmonicity, rather than as a harmonic spectrum (which is the preferred scientific description). However, the term "harmonic" may occasionally be used in the spectral sense as a contrasting term to "inharmonic".

A second problem arises because a spectrum in motion may still preserve this simple relationship between its constituent partials, as it moves. To put it simply, a portamento is pitched in the spectral sense. It is difficult to speak of a portamento as "having a pitch" in the sense of conventional harmony. This sense of "having a pitch" i.e. being able to assign a pitch to a specific class like E-flat or C#2 is quite a different concept from the timbral concept of pitch described here. We will therefore refer to that traditional concept as *Hpitch*, an abbreviation for pitch-as-related-to-conventional-harmony.

Perception of Hpitch depends on the existence of a frame of reference (see Chapter 1). Even with steady (non-portamento) pitches, we may still have no sense of Hpitch if the pitches are selected at random from the continuum of values, though often our cultural predispositions cause us to "force" the notes onto our preconceived notion of where they "ought" to be. In the sound example we hear first a set of truly random pitches, next a set of pitches on a HArmonic field, then a set of pitches approximable to a HArmonic field and finally the 'same' set locked onto that HArmonic field. **(Sound example 2.3)**.

We can, in fact, generate a perception of a single, if unstable, pitch from an event containing very many different pitches scattered over a narrow band. The narrower the band, the more clearly is a single pitch defined. **(Sound example 2.4)**.

Once we confine our selection of pitches to a given reference frame (scale, mode, HArmonic field) we establish a clear sense of Hpitch for each event.

Returning now to portamenti and considering rising portamenti, it is possible to relate such moving pitches to Hpitch if the portamenti have special properties. Thus, if they are onset-weighted and the onsets located in a fixed HArmonic field, we will assign Hpitch to the events, regarding the portamenti as ornaments or articulations to those Hpitches. **(Sound example 2.5a)**. Similarly, end-focused portamenti (often heard in popular music singing styles) where the portamenti settle clearly onto values in a HArmonic field, will be perceived as anacrusial ornaments to Hpitches. **(Sound example 2.5b)**. Portamenti with no such weighting, however, will not be perceived to be Hpitched. **(Sound example 2.5c)**.

The same arguments apply to falling portamenti and even more powerfully, to complexly moving portamenti. **(Sound example 2.6)**. Nevertheless all these sounds are pitched in the spectral sense.

Using our new computer instruments it becomes possible to follow and extract the pitch from a sound event, but this does not necessarily (or usually) mean that we are assigning an Hpitch, or a set of Hpitches to it. The pitch flow of a speech stream can be followed, extracted and applied to an entirely different sound object, establishing a definite relationship between the two without any conception of Hpitch or scale, mode or HArmonic field entering into our thinking, *or our perception*. This should be borne in mind while reading this Chapter as it is all too easy for musicians to slide imperceptibly into thinking of pitch as Hpitch!

A good example in traditional musical practice of pitch not treated as Hpitch can be found in Xenakis' *Pithoprakta* where portamenti are organised in statistical fields governed by Poisson's formula or in portamenti of portamenti, which themselves rise and fall without having any definite Hpitch. (See *On Sonic Art*). **(Sound example 2.7)**.

In contrast, Paul de Marinis uses pitch extraction on natural speech recordings, subsequently reinforcing the speech pitches on synthetic instruments so the speech appears to sing. **(Sound example 2.8)**.

PITCH-TRACKING

It seems odd so early in this book to tackle what is perhaps one of the most difficult problems in instrument design. Whole treatises have been written on pitch-detection and its difficulties. The main problem for the instrument designer is that the human ear is remarkably good at pitch detection and even the best computing instruments do not quite match up to it. This being said, we can make a reasonably good attempt in most cases.

Working in the time domain, we will recognise pitch if we can find a cycle of the waveform (a wavelength) and then correlate that with similar cycles immediately following it. The pitch is then simply one divided by the wavelength. This is *pitch-tracking by auto-correlation* **(Appendix p70)**.

We can also attempt *pitch–tracking by partial analysis* (**Appendix p71**). Hence in the frequency domain we would expect to have a sequence of (time) windows, in which the most significant frequency information has been extracted in a number of channels (as in the *phase vocoder*). Provided we have many more channels than there are partials in the sound, we will expect that the partials of the sound have been separated into distinct channels. We must then separate the true partials from the other information by looking for the peaks in the data.

Then, as our previous discussion indicated, we must find the highest common factor of the frequencies of our partials, which will exist if our instantaneous spectrum is harmonic. Unfortunately, if we allow sufficiently small numbers to be used, then, within a given limit of accuracy, any set of partial frequencies values will have a highest common factor. e.g. partials at 100.3, 201, 307 and 513.5 have an HCF of 0.1. We must therefore reject absurdly small values. A good lower limit would be 16 cycles, the approximate lower limit of pitch perception in humans.

The problem of pitch–tracking by partial analysis can in fact be simplified if we begin our search on a quarter–tone grid, and also if we know in advance what the spectral content of the source sound is (see **Appendix p71**). In such relatively straightforward cases pitch–tracking can be very accurate, with perhaps occasional octaviation problems (the Hpitch can be assigned to the wrong octave). However, in the general case (e.g. speech, or synthesised sequences involving inharmonic spectra), where we wish to track pitch independently of a reference frame, and where we cannot be sure whether the incoming sound will be pitched or not, the problem of pitch–tracking is hard.

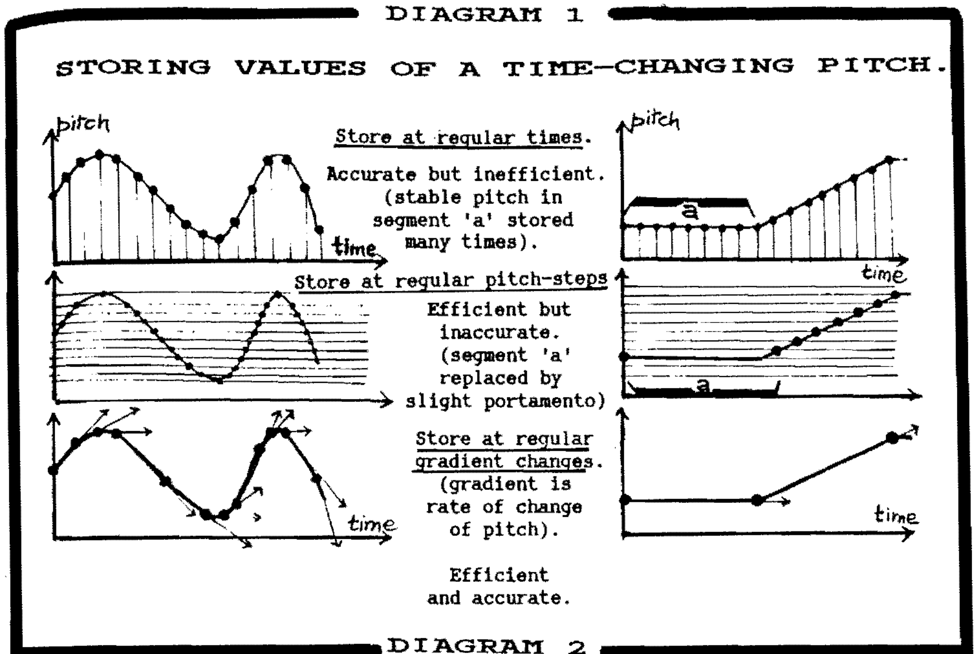
For even greater certainty we might try correlating the data from the time–domain with that from the frequency domain to come up with our most definitive solution.

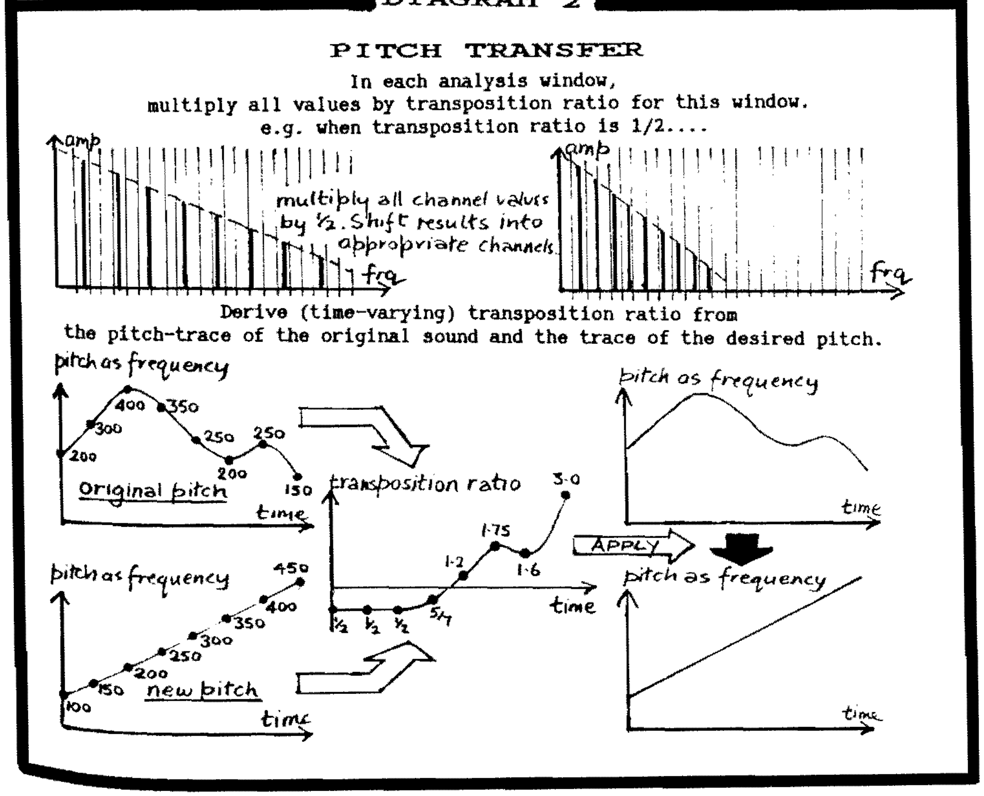
The pitch data might be stored in (the equivalent of) a *breakpoint table* of time/frequency values. In this case we need to decide upon the frequency resolution of our data, i.e. how much must a pitch vary before we record a new value in our table? More precisely, if a pitch is changing, when is the rate of change adjudged to have changed? (See **Diagram 1**).

If we are working on an Hpitch reference frame the task is, of course, much simpler. If we do not confine ourselves to such frames, to be completely rigorous we could store the pitch value found at every window in the frequency domain representation, But this is needlessly wasteful. Better to decide on our own ability to discriminate rates of pitch motion and to give the pitch–detection instrument a portamento–rate–change threshold which, when exceeded, causes a new value to be recorded in our pitch data file.

PITCH TRANSFER

Once we have established a satisfactory pitch–trace for a sound, we can modify the pitch of the original sound and this is most easily considered in the frequency domain. We can provide a new (changing) pitch–trace, either directly, or from a second sound. By comparing the two traces, a pitch–following instrument will deduce the ratio between the new pitch and the original pitch at a particular window time (the instantaneous transposition ratio), then multiply the frequency values in each channel of the original window by that ratio, hence altering the perceived pitch in the resynthesized sound. (See **Diagram 2**).





There are also some fast techniques that can be used in special cases. Once the pitch of a sound is known, we can transpose it up an octave using *comb-filter transposition* (**Appendix p65**). Here we *delay* the signal by half the wavelength of the fundamental and add the result to the original signal. This process cancels (by *phase-inversion*) the odd harmonics while reinforcing the even harmonics. Thus if we start with a spectrum whose partials are at 100, 200, 300, 400, 500 etc with a fundamental at 100Hz, we are left with partials at 200, 400 etc whose fundamental lies at 200Hz, an octave above the original sound. A modification of the technique, using the *Hilbert transform*, allows us to make an octave downward transposition in a similar manner. The process is particularly useful because it does not disturb the contour of the spectrum (the formants are not affected : see Chapter 3) so it can be applied successfully to vocal sounds.

It is important to emphasize that pitch manipulation does not have to be embedded in a traditional approach to Hpitches. The power of pitch-tracking is that it allows us to trace and transfer the most subtle or complex pitch flows and fluctuations without necessarily being able to assign specific Hpitch values at any point. For example, the subtleties of portamento and vibrato articulation in a particular established vocal or instrumental idiom, or in a naturally recorded bird or animal cry, could be transferred to an arbitrarily chosen non-instrumental, non-vocal, or even synthetic sound-object. It would even be possible to interpolate between such articulation styles without at any stage having a quantifiable (measurable) or notatable representation of them - we do not need to be able to measure or analytically explain a phenomenon to make an aesthetic decision about it.

NEW PROBLEMS IN CHANGING PITCH

Changing the pitch of a musical event would seem, from a traditional perspective, to be the most obvious thing to do. Instruments are set up, either as collections of similar objects (strings, metal bars, wooden bars etc.), or with variable access to the same objects (flute fingerholes, violin fingerboard, brass valves) to permit similar sounds with different pitches to be produced rapidly and easily.

There are two problems when we try to transfer this familiar notion to sound composition. Firstly, we do not necessarily want to confine ourselves to a finite set of pitches or to steady pitches (an Hpitch set). More importantly the majority of sounds do not come so easily prepackaged either because the circumstances of their production cannot be reproduced (breaking a sheet of glass ... every sheet will break differently, no matter how much care we take!), or because the details of their production cannot be precisely remembered (a spoken phrase can be repeated at different pitches by the same voice within a narrow range but, assuming natural speech inflection, the fine details of articulation cannot usually be precisely remembered).

In fact changing the pitch on an instrument does involve some spectral compromises, e.g. low and high pitches on the piano have a very different spectral quality but we have come to regard these discrepancies as acceptable through the skill of instrument design (the piano strings resonate in the same sound-box and there is a relatively smooth transition in quality from low to high strings), and the familiarity of tradition. We are not, in fact, changing the pitch of the original sound, but producing another sound, whose relationship to the original is acceptable.

The ideal way, therefore, to change the pitch of the sound is to build a synthetic model of the sound, then alter its fundamental frequency. However this is a very complicated task, and adopting this approach for every sound we use, would make sound composition unbelievable arduous. So we must find alternative approaches.

DIFFERENT APPROACHES TO PITCH-CHANGING

In the time-domain, the obvious way to change the pitch is to change the wavelength of the sound. In classical tape studios the only way to do this was to speed up (or slow down) the tape. On the computer, we simply re-read the digital data at a different step (instead of every one sample, read every 2 samples, or every 1.3 samples). This is *tape-speed variation*. This automatically makes every wavelength shorter (or longer) and changes the pitch. Unfortunately, it also makes the source sound shorter (longer). If this doesn't matter, it's the simplest method to adopt, but with segmented sounds (speech, melodies) or moving sounds (e.g. portamenti) it changes their perceived speed. (**Sound example 2.9**).

Computer control makes such "tape-speed" variation a more powerful tool as we can precisely control the speed change trajectory or specify a speed-change in terms of its final velocity (*tape acceleration*). The availability of time-variable processing gives us a new order of compositional control of such time-varying processes. (**Sound example 2.10**).

Waveset transposition is an unconventional approach which avoids the time-distortion involved in tape-speed variation and can be used for integral multiples of the frequency. Here, each waveset (in the sense of a pair of zero-crossings : **Appendix p50**) is replaced by N shortened copies occupying the same time as the original one. (**Diagram 3** and **Appendix p51**). This technique is very fast to compute but often introduces strange, signal dependent (i.e. varying with the signal) artefacts. (It can therefore be used as a process of *constructive distortion* in its own right!) Grouping the wavesets in pairs, triplets etc., before reproducing them, can affect the integrity of reproduction of the sound at the new pitch (see **Diagram 4**). The grouping to choose again depends on the signal.

With accurate pitch-tracking this technique can be applied to true wavecycles (deduced from a knowledge of both the true wavelength and the zero-crossing information) and should avoid producing artefacts. (**Diagram 5**).

The technique can also be used to transpose the sound downwards, replacing N wavesets or wavecycles by just one of them, enlarged, but too much information may be lost (especially in a complex sound) to give a satisfactory result in terms of just pitch- shifting. (**Appendix p51** : **Sound example 2.11a**).

A more satisfactory time-domain approach is through *brassage*. To lower the pitch of the sound, we *cut* the sound into short segments, slow them down as in tape-speed variation, which lengthens them, then *splice* them together again so they overlap sufficiently to retain the original duration. It is crucial to use segments in the grain time-frame (see Chapters 1 & 4), so that each segment is long enough to carry instantaneous pitch information, but not long enough to have a perceptible internal structure which would lead to unwanted echo effects within the pitch-changed sound. This technique, used in the *harmoniser*, works quite well over a range of one octave, up or down, but beyond this, begins to introduce significant artefacts: the signal is transformed as well as pitch-shifted. (**Sound example 2.12**).

DIAGRAM 3

**Replace each waveset by three shortened copies of itself.
Wavelength redduced to 1/3, therefore frequency X 3,
hence transpose up by interval of a 12th.**

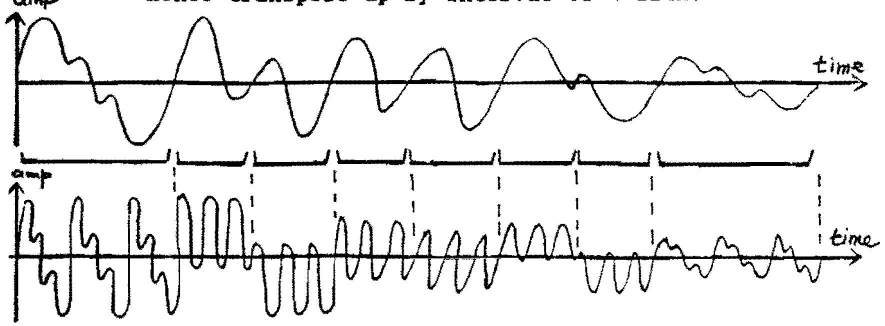


DIAGRAM 4

Take wavesets in groups of three.
Replace each 3-set by 3 shortened copies of itself.

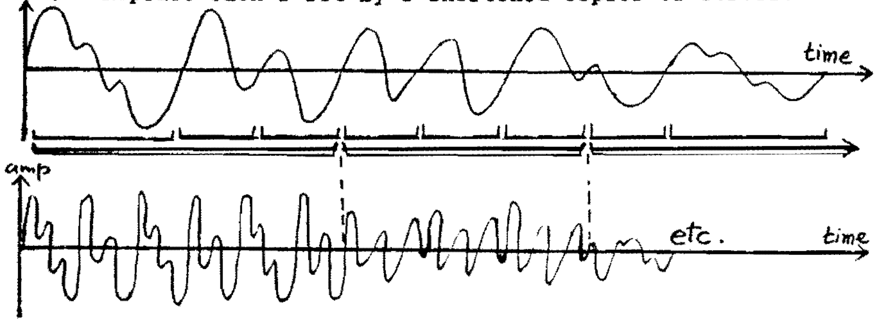
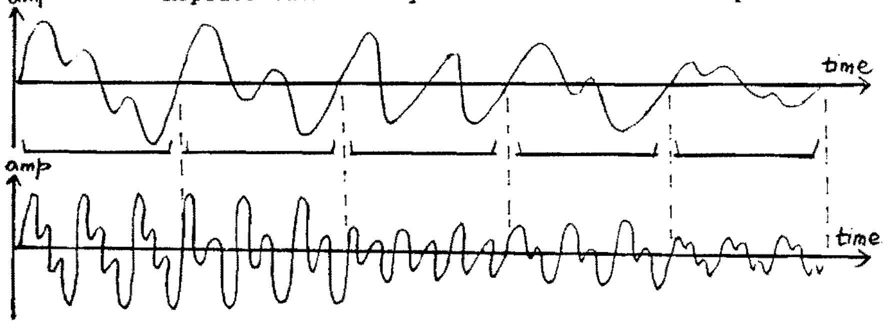


DIAGRAM 5

**Find true wavecycles, with help of a pitch-tracking instrument.
Replace each wavecycle with 3 shortened copies.**



In the frequency domain, pitch-shifting is straightforward. We need only multiply the frequencies of the components in each channel (in each window) by an appropriate figure, the transposition ratio. As this does not change the window duration, the pitch is shifted without changing the sounds duration. This is *spectral shifting*. **(Sound example 2.13)**.

PRESERVING THE SPECTRAL CONTOUR

All these approaches, however, shift the formant-characteristics of the spectrum. The problem here is that certain spectral characteristics of a sound are determined by the overall shape of the spectrum at each moment in time (the spectral contour) and particularly by peaks in the spectral contour known as formants (see Chapter 3). Thus the vowel sound "a" will be found to be related to various peaks in the spectral contour. If we change the pitch at which "a" is sung, the partials in the sound will all move up (or down) the frequency ladder. However, the spectral peaks will *remain where they were* in the frequency space. Thus, if there was a peak at around 3000 Hz, we will continue to find a peak at around 3000 Hz. (See **Appendix p10**).

Simply multiplying the channel frequencies by a transposition ratio causes the whole spectrum, and hence the spectral peaks (formants), to move up (or down) the frequency space. Hence the formants are moved and the "a"-ness of the sound destroyed. **(Appendix p16)**.

A more sophisticated approach therefore involves determining the spectral contour in each window, retaining it and then superimposing the unshifted contour on the newly shifted partials. The four stages might be as follows...

(a) Extract the spectral contour using *linear predictive coding* (LPC) **(Appendix pp12-13)**.

(b) Extract the partials with the *phase vocoder*. **(Appendix p11)** or with a fine-grained LPC (see *spectral-focusing* below).

(c) Flatten the spectrum using the inverse of the spectral contour.

(d) Change the spectrum.

(e) Reimpose the original spectral contour.(See **Appendix p17**).

Ideally this approach of separating the formant data and the partial data should be applied even when merely imposing vibrato on a sound (see Chapter 10) but it is computationally intensive and, except in the case of the human voice, probably excessively fastidious in most situations.

Formant drift is an obvious problem when dealing with speech sounds, but needs to be borne in mind more generally. An instrument is characterised often by a single soundbox (piano, violin) which provides a relatively fixed background spectral contour for the entire gamut of notes played on it. We are, however, more obviously aware of formant drift in situations (like speech) where the *articulation* of formants is significant.

DUAL PITCH

Various compositional processes allow us to generate more than one pitch from a sound. For example, using the *harmoniser* approach we can shift the pitch of a sound without altering its duration and then *mix* this with the original pitch. Apart from the fact that we can apply this technique to any sound, it differs from simply playing two notes on the same instrument because the variations and microfluctuations of the two pitches remain fairly well in step, a situation impossible to achieve with two separate performers, though it may be closely approached by a single performer using e.g. double-stopping on a stringed instrument. The technique tends to be more musically interesting when used on subtly fluctuating sounds, rather than as a cost-saving way of adding conventional HArmony to a melodic line. (**Sound example 2.14**).

As usual, the *harmoniser* algorithm introduces significant artefacts over larger interval shifts. An alternative approach, therefore, is to use *spectral shifting* in the frequency domain, superimposing the result on the original source. In fact, we can use this kind of spectral shifting to literally split the spectrum in two, shifting only a part of the spectrum. The two sets of partials thus generated will imply two different fundamentals and the sound will appear to have a split pitch. (**Appendix p18**). (**Sound example 2.15**).

All these techniques cane be applied *dynamically* so that the pitch of a sound *gradually* splits in two. (**Sound example 2.16**).

Small pitch-shifts, superimposed on the original sound add "body" to the sound, producing the well known "chorus" effect (an effect produced naturally by a chorus of similar singers, or similar instruments playing the same pitches. where small variations in tuning between individual singers, or players, broaden the spectral band of the resultant massed sound). (**Sound example 2.17**).

A different kind of pitch duality can be produced when the focus of energy in the spectrum (spectral peak) moves markedly above a fixed pitch or over a pitch which is moving in a contrary direction. There are not truly two pitches present in these cases but percepts of conflicting frequency motions within the sound can certainly be established. With very careful control, including the phasing in and out of partials at the top and bottom of the spectrum, sounds can be created which get higher in Hpitch yet lower in tessitura (or lower in Hpitch but higher in tessitura) – the so called *Shepard Tones*. (see *On Sonic Art* and **Appendix p72**). (**Sound example 2.18**).

Similarly, by appropriate filtering, we can individually reinforce the harmonics (or partials) in a spectrum so that our attention is drawn to them as perceived pitches in their own right (as in Tibetan chanting or Tuvan harmonics singing). (**Sound example 2.19**).

Another pitch phenomena which is worth noting is the pitch drift associated with *spectral stretching* (see **Appendix p19** and Chapter 3). If the partials of a pitch-spectrum are gradually moved so the relationship ceases to be harmonic (no longer whole number multiples of the fundamental), the sound will begin to present several pitches to our perception (like bell sounds). Moreover if the stretch is upwards, even if the fundamental frequency is present in the spectrum and remains unchanged, the lowest perceived pitch may gradually move upwards. (**Sound example 2.20**).

TESSITURA CHANGE OF UNPITCHED SOUNDS

The techniques of pitch change we have described can be applied to sounds without any definite pitch. Inharmonic sound will, in general, be transposed in a similar way to pitched sounds. *Waveset transposition* will usually change the centre-of-energy (the "pitch-band" or tessitura) of noisy sounds so they appear to move higher (or lower). (**Sound example 2.21**). The *harmoniser* will usually have the same effect. Splitting the spectrum of a broad-band, noise-based sound using *spectral shifting* may not have any noticeable perceptual effect on the sound, even when the split is quite radical and *chorusing* will often be unnoticeable as the spectrum is already full of energy. However, the problem of formant shifting when transposing will apply equally well to, e.g. unvoiced speech sounds. (**Sound example 2.22**).

We can also use this technique to give a sense of pitch motion in unpitched sounds – noise portamenti. (**Sound example 2.23**).

PITCH CREATION

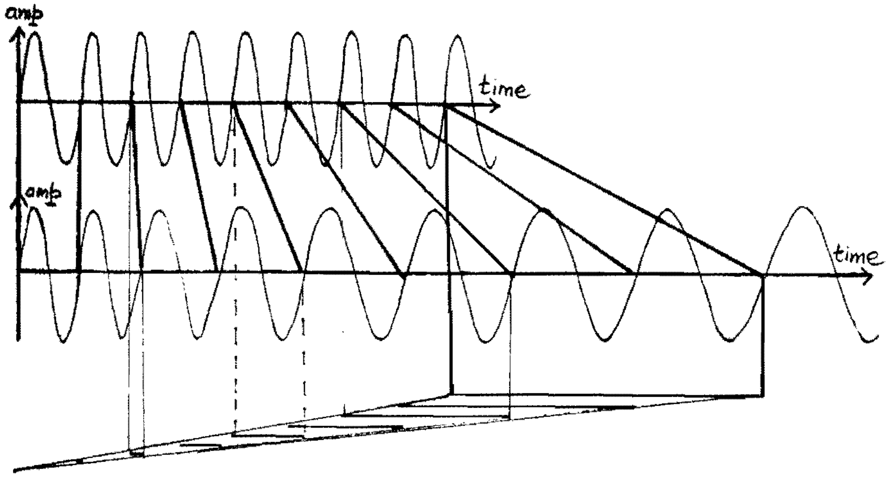
It is possible to give pitch qualities to initially unpitched sounds. There are two approaches to this which, at a deeper level, are very similar. A filter is an instrument which reinforces or suppresses particular parts of the spectrum. A filter may work over a large band of frequencies (when it is said to have a small Q) or over a narrow band of frequencies (when it has a large Q). We can use a *filter* not only to remove or attenuate parts of the spectrum but, by inversion, to accentuate what remains (*bandpass filter* : **Appendix p7**). In particular, narrow filters with very high "Q" (so that the bands allowed to pass are very abruptly defined), can single out narrow ranges of partials from a spectrum. A very narrow fixed filter can thus impose a marked spectral peak on the noisiest of sounds, giving it a pitched quality.

In fact this process works best on noisier sounds because here the energy will be distributed over the entire spectrum and wherever we place our filter band(s), we will be sure to find some signal components to work on. Very tight Q on the filter will produce oscillator-type pitches, whilst less tight Q and broader bands will produce a vaguer focusing of the spectral energy. There are many degrees of "pitchiness" between pure noise and a ringing oscillator. (**Sound example 2.24**). We can also force our sound through a stack of filters, called a *filter bank*, producing "chords", and, increasing Q with time, move from noise towards such chords. (**Sound example 2.25**).

In a sound with a simpler spectrum, a narrow, tight filter may simply miss any significant spectral elements – we may end up with silence!

A very sophisticated approach to this process is *spectral focusing* (**Appendix p20**). In this process, we first make an analysis of a sound with (possibly time-varying) pitch using *Linear Predictive Coding* (**Appendix pp12-13**). However, we vary the size of the analysis window through time. With a normal size analysis window we will extract the spectral contour (the formants) (see above). With a very fine analysis window, however, we will pick out the individual partials.

DIAGRAM 6



Time varying delay between two sounds,
one of which is a time-varying time-stretched version of the other.

If we now use the analysis data as a set of (time-varying) filters on an input noise sound, wherever the analysis window was normal-sized the resultant filters will impose the formant characteristics of the original sound on the noise source (e.g. analysed voiced speech will produce unvoiced speech),but where the window size was very fine, we will have generated a set of very narrow-Q filters at the (time-varying) frequencies of the original partials. These will then act on the noise to produce something very close to the original signal.

If the original analysis window-size varied in time from normal to fine, our output sound would vary from formant-shaped noise to strongly pitched sound (e.g. from an analysis of pitched speech, our new sound would pass from unvoiced to voiced speech). This then provides a sophisticated means to pass from noise to pitch in a complexly evolving sound-source.

The second approach to pitch-generation is to use *delay*. As digital signals remain precisely in time, the delay between equivalent samples in the original and delayed sound will remain exactly fixed. If this delay is short enough, we will hear a pitch corresponding to one divided by delay time, whatever sound we input to the system. This technique is known as *comb filtering*. Longer delays will give lower and less well defined pitches. **(Sound example 2.26)**.

Both these techniques allow us to produce dual-pitch percepts with the pitch of the source material moving in some direction and the pitch produced by the delay or filtering fixed, or moving in a different sense (with time-variable filtering or delay).

Producing portamenti is an even simpler process. When a sound is *mixed* with a very slightly *time-stretched* (or shrunk) copy of itself, we will produce a gradually changing delay (See **Diagram 6**). If the sounds are start-synchronised, this will produce a downward portamento. If the sounds are end-synchronised, we will produce an upward portamento. We may work with more than two time-varied copies. **(Sound example 2.27)**.

Phasing or Flangeing, often used in popular music, relies on such delay effects. In this case the signal is delayed by different amounts in different frequency registers using an *all-pass filter* **(Appendix p9)** and this shifted signal is allowed to interact with the unchanged source.

The production of pitch-motion seems an appropriate place to end this Chapter as it stresses once again the difference between pitch and Hpitch and the power of the new compositional tools to provide control over pitch-in-motion.

CHAPTER 3

SPECTRUM

WHAT IS TIMBRE ?

The spectral characteristics of sounds have, for so long, been inaccessible to the composer that we have become accustomed to lumping together all aspects of the spectral structure under the catch-all term "timbre" and regarding it as an elementary, if unquantifiable, property of sounds. Most musicians with a traditional background almost equate "timbre" with instrument type (some instruments producing a variety of "timbres" e.g. pizz, arco, legno, etc). Similarly, in the earliest analogue studios, composers came into contact with oscillators producing featureless pitches, noise generators, producing featureless noise bands, and "envelope generators" which added simple loudness trajectories to these elementary sources. This gave no insight into the subtlety and multidimensionality of sound spectra.

However, a whole book could be devoted to the spectral characteristics of sounds. The most important feature to note is that all sound spectra of musical interest are time-varying, either in micro-articulation or large-scale motion.

HARMONICITY - INHARMONICITY

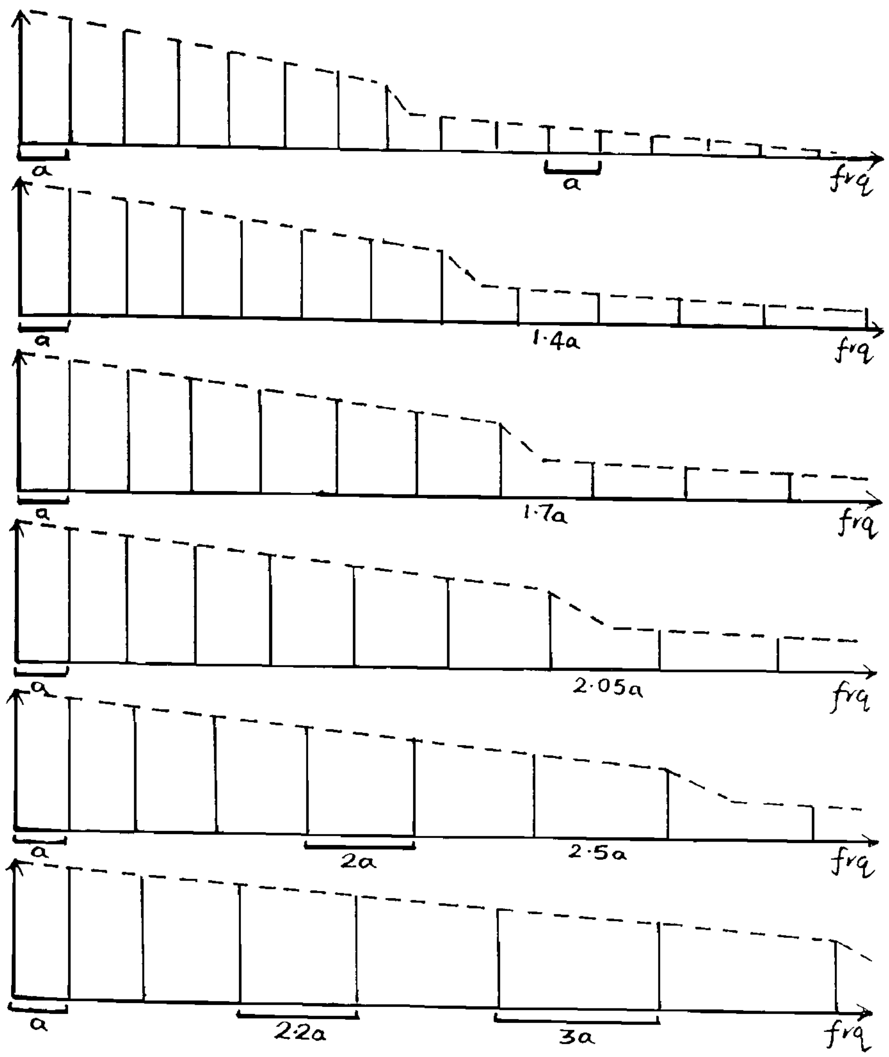
As discussed in Chapter 2, if the partials which make up a sound have frequencies which are exact multiples of some frequency in the audible range (known as the fundamental) and, provided this relationship persists for at least a grain-size time-frame, the spectrum fuses and we hear a specific (possibly gliding) pitch. If the partials are not in this relationship, and provided the relationships (from window to window) remain relatively stable, the ear's attempts to extract harmonicity (whole number) relationships amongst the partials will result in our hearing several pitches in the sound. These several pitches will trace out the same micro-articulations and hence will be fused into a single percept (as in a bell sound). The one exception to this is that certain partials may decay more quickly than others without destroying this perceived fusion (as in sustained acoustic bell sounds).

In **Sound example 3.1** we hear the syllable "ko->u" being gradually *spectrally stretched* : **Appendix p19**). This means that the partials are moved upwards in such a way that their whole number relationships are preserved less and less exactly and eventually lost. (See **Diagram 1**). Initially, the sound appears to have an indefinable "aura" around it, akin to phasing, but gradually becomes more and more bell-like.

It is important to understand that this transformation "works" due to a number of factors apart from the harmonic\inharmonic transition. As the process proceeds, the tail of the sound is gradually *time-stretched* to give it the longer decay time we would expect from an acoustic bell. More importantly, the morphology (changing shape) of the spectrum is already bell-like. The syllable "ko->u" begins with a very short broad band spectrum with lots of high-frequency information ("k") corresponding to the initial clang of a bell. This leads immediately into a steady pitch, but the vowel formant is varied from "o" to "u", a process which gradually fades out the higher partials leaving the lower to continue. Bell sounds have this similar property, the lower partials, and hence the lower heard pitches, persisting longer than the higher components. A different initial morphology would have produced a less bell-like result.

This example (used in the composition of *Vox 5*) illustrates the importance of the time-varying structure of the spectrum (not simply its loudness trajectory).

DIAGRAM 1



We may vary this *spectral stretching* process by changing the overall stretch (i.e. the top of the spectrum moves further up or further down from its initial position) and we may vary the type of stretching involved. **(Appendix p19). (Sound example 3.2).**

Different types of stretching will produce different relationships between the pitches heard within the sounds.

Note that, small stretches produce an ambiguous area in which the original sound appears " coloured" in some way rather than genuinely multi-pitched. **(Sound example 3.3).** Inharmonicity does not therefore necessarily mean multipitchedness. Nor (as we have seen from the "ko->u" example), does it mean bell sounds. Very short inharmonic sounds will sound percussive, like drums, strangely coloured drums, or akin to wood-blocks **(Sound example 3.4).** These inharmonic sounds can be transposed and caused to move (subtle or complex pitch-gliding) just like pitched sounds (also see Chapter 5 on *Continuation*).

Proceeding further, the spectrum can be made to vary, either slowly or quickly, between the harmonic and the inharmonic creating a dynamic interpolation between a harmonic and an inharmonic state (or between any state and something more inharmonic) so that a sound changes its spectral character as it unfolds. We can also imagine a kind of harmonic to inharmonic vibrato-like fluctuation within a sound. **(Sound example 3.5).**

Once we vary the spectrum too quickly, and especially if we do so irregularly, we no longer perceive individual moments or grains with specific spectral qualities. We reach the area of noise (see below).

When transforming the harmonicity of the spectrum, we run into problems about the position of formants akin to those encountered when pitch-changing (see Chapter 2) and to preserve the formant characteristics of the source we need to preserve the spectral contour of the source and apply it to the resulting spectrum (see *formant preserving spectral manipulation* : **Appendix p17**).

FORMANT STRUCTURE

In any window, the contour of the spectrum will have peaks and troughs. The peaks, known as formants, are responsible for such features as the vowel-state of a sung note. For a vowel to persist, the spectral contour (and therefore the position of the peaks and troughs) must remain where it is even if the partials themselves move. (See **Appendix p10**).

As we know from singing, and as we can deduce from this diagram, the frequencies of the partials in the spectrum (determining pitch(es), harmonicity-inharmonicity, noisiness) and the position of the spectral peaks, can be varied independently of each other. This is why we can produce coherent speech while singing or whispering. **(Sound example 3.6).**

Because most conventional acoustic instruments have no articulate time-varying control over spectral contour (one of the few examples is hand manipulable brass mutes), the concept of formant control is less familiar as a musical concept to traditional composers. However, we all use articulate formant control when speaking.

It is possible to extract the (time varying) spectral contour from one signal and impose it on another, a process originally developed in the analogue studios and known as *vocoding* (*no* connection with the *phase vocoder*). For this to work effectively, the sound to be vocoded must have energy distributed over the whole spectrum so that the spectral contour to be imposed has something to work on. Vocoding hence works well on noisy sounds (e.g. the sea) or on sounds which are artificially prewhitened by adding broad band noise, or subjected to some noise producing distortion process. **(Sound example 3.7).**

It is also possible to normalise the spectrum before imposing the new contour. This process is described in Chapter 2, and in the under *formant preserving spectral manipulation* in **Appendix p17.**

Formant-variation of the spectrum does not need to be speech-related and, in complex signals, is often more significant than spectral change. We can use *spectral freezing* to freeze certain aspects of the spectrum at a particular moment. We hold the frequencies of the partials, allowing their loudnesses to vary as originally. Or we can hold their amplitudes stationary, allowing the frequencies to vary as originally. In a complex signal, it is often holding steady the amplitudes, and hence the spectral contour, which produces a sense of "freezing" the spectrum when we might have anticipated that holding the frequencies would create this percept more directly. **(Sound example 3.8).**

NOISE, "NOISY NOISE" & COMPLEX SPECTRA

Once the spectrum begins to change so rapidly and irregularly that we cannot perceive the spectral quality of any particular grain, we hear "noise". Noise spectra are not, however, a uniform grey area of musical options (or even a few shades of pink and blue) which the name (and past experience with noise generators) might suggest. The subtle differences between unvoiced staccato "t", "d", "p", "k", "s", "sh", "f", the variety amongst cymbals and unpitched gongs give the lie to this.

Noisiness can be a matter of degree, particularly as the number of heard out components in an inharmonic spectrum increases gradually to the point of noise saturation. It can, of course, vary formant-wise in time: whispered speech is the ideal example. It can be more or less focused towards static or moving pitches, using *band-pass filters* or *delay* (see Chapter 2), and it can have its own complex internal structure. In **Sound example 3.9** we hear portamentoing inharmonic spectra created by filtering noise. This filtering is gradually removed and the bands become more noise-like.

A good example of the complexity of noise itself is "noisy noise", the type of crackling signal one gets from very poor radio reception tuned to no particular station, from masses of broad-band click-like sounds (either in regular layers - cicadas - or irregular - masses of breaking twigs or pebbles falling onto tiles - or semi-regular - the gritty vocal sounds produced by water between the tongue and palate in e.g. Dutch "gh") or from extremely time-contracted speech streams. There are also fluid noises produced by portamentoing components, e.g. the sound of water falling in a wide stream around many small rocks. These shade off into the area of "Texture" which we will discuss in Chapter 8. **(Sound example 3.10).**

These examples illustrate that the rather dull sounding word "noise" hides whole worlds of rich sonic material largely unexplored in detail by composers in the past.

Two processes are worth mentioning in this respect. Noise with transient pitch content like water falling in a stream (rather than dripping, flowing or bubbling), might be pitch–enhanced by *spectral tracing* (see below). (**Sound example 3.11**). Conversely, all sounds can be amassed to create a sound with a noise–spectrum if superimposed randomly in a sufficiently frequency–dense and time–dense way. At the end of **Sound example 3.9** the noise band finally resolves into the sound of voices. The noise band was in fact simply a very dense superimposition of many vocal sounds.

Different sounds (with or without harmonicity, soft or hard–edged, spectrally bright or dull, grain–like, sustained, evolving, iterated or sequenced) may produce different qualities of noise (see Chapter 8 on Texture). There are also undoubtedly vast areas to be explored at the boundaries of inharmonicity/noise and time–fluctuating–spectrum/noise. (**Sound example 3.12**).

A fruitful approach to this territory might be through *spectral focusing*, described in Chapter 2 (and **Appendix p20**). This allows us to extract, from a pitched sound, either the spectral contour *only*, or the true partials, and to then use this data to filter a noise source. The filtered result can vary from articulated noise formants (like unvoiced speech) following just the formant articulation of the original source, to a reconstitution of the partials of the the original sound (and hence of the original sound itself). We can also move fluidly between these two states by varying the analysis window size through time. This technique can be applied to any source, whether it be spectrally pitched (harmonic), or inharmonic, and gives us a means of passing from articulate noise to articulate not–noise spectra in a seamless fashion.

Many of the sound phenomena we have discussed in this section are complex concatenations of simpler units. It is therefore worthwhile to note that any arbitrary collection of sounds, especially *mixed* in mono, has a well–defined time–varying spectrum – a massed group of talkers at a party; a whole orchestra individually, but simultaneously, practising their difficult passages before a concert. At each moment there is a composite spectrum for these events and any portion of it could be grist for the mill of sound composition.

SPECTRAL ENHANCEMENT

The already existing structure of a spectrum can be utilised to enhance the original sound. This is particularly important with respect to the onset portion of a sound and we will leave discussion of this until Chapter 4. We may reinforce the total spectral structure, adding additional partials by *spectral shifting* the sound (without changing its duration) (**Appendix p18**) and *mixing* the shifted spectrum on the original. As the digital signal will retain its duration precisely, all the components in the shifted signal will line up precisely with their non–shifted sources and the spectrum will be thickened while retaining its (fused) integrity. Octave enhancement is the most obvious approach but any interval of transposition (e.g. the tritone) might be chosen. The process might be repeated and the relative balance of the components adjusted as desired. (**Appendix p48**). (**Sound example 3.13**).

A further enrichment may be achieved by mixing an already stereo spectrum with a pitch–shifted version which is left–right inverted. Theoretically this produces merely a stage–centre resultant spectrum but in practice there appear to be frequency dependent effects which lend the resultant sound a new and richer spatial "fullness". (**Sound example 3.14**).

Finally, we can introduce a sense of multiple–sourcedness (!) to a sound (e.g. make a single voice appear crowd–like) by adding small random time–changing perturbations to the loudnesses of the spectral components (*spectral shaking*).This mimics part of the effect of several voices attempting to deliver the same information. (**Sound example 3.15**). We may also perturb the partial frequencies (**Sound example 3.16**).

SPECTRAL BANDING

Once we understand that a spectrum contains many separate components, we can imagine processing the sound to isolate or separate these components. *Filters*, by permitting components in some frequency bands to pass and rejecting others, allow us to select parts of the spectrum for closer observation. With dense or complex spectra the results of filtering can be relatively unexpected revealing aspects of the sound material not previously appreciated. A not–too–narrow and static *band pass filter* will transform a complex sound–source (usually) retaining its morphology (time–varying shape) so that the resulting sound will relate to the source sound through its articulation in time. (**Sound example 3.17**).

A *filter* may also be used to isolate some static or moving feature of a sound. In a crude way, filters may be used to eliminate unwanted noise or hums in recorded sounds, especially as digital filters can be very precisely tuned. In the frequency domain, spectral components can be eliminated on a channel–by–channel basis, either in terms of their frequency location (using *spectral splitting* to define a frequency band and setting the band loudness to zero) or in terms of their timevarying relative loudness (*spectral tracing* will eliminate the N least significant, i.e. quietest, channel components, window by window. At an elementary level this can be used for signal–dependent noise reduction. But see also "Spectral Fission" below). More radically, sets of narrow *band pass filters* can be used to force a complex spectrum onto any desired Hpitch set (HArmonic field in the traditional sense). (**Sound example 3.18**).

In a more signal sensitive sense a filter or a frequency–domain channel selector can be used to separate some desired feature of a sound, e.g. a moving high frequency component in the onset, a particular strong middle partial etc, for further compositional development. In particular, we can separate the spectrum into parts (using *band pass filters* or *spectral splitting*) and apply processes to the N separated parts (e.g. pitch–shift, add vibrato) and then recombine the two parts perhaps reconstituting the spectrum in a new form. However, if the spectral parts are changed too radically e.g. adding completely different vibrato to each part, they will not fuse when *remixed*, but we may be interested in the gradual dissociation of the spectrum. This leads us into the next area.

Ultimately we may use a procedure which follows the partials themselves, separating the signal into its component partials (*partial tracking*). This is quite a complex task which will involve *pitch tracking* and pattern–matching (to estimate where the partials might lie) on a window by window basis. Ideally it must deal in some way with inharmonic sounds (where the form of the spectrum is not known in advance) and noise sources (where there are, in effect, no partials). This technique is however particularly powerful as it allows us to set up an *additive synthesis* model of our analysed sound and thereby provides a bridge between unique recorded sound–events and the control available through synthesis.

SPECTRAL FISSION & CONSTRUCTIVE DISTORTION

We have mentioned several times the idea of spectral fusion where the parallel micro-articulation of the many components of a spectrum causes us to perceive it as a unified entity – in the case of a harmonic spectrum, as a single pitch. The opposite process, whereby the spectral components seem to split apart, we will describe as spectral fission. Adding two different sets of vibrato to two different groups of partials within the same spectrum will cause the two sets of partials to be perceived independently – the single aural stream will split into two. (**Sound example 3.19**).

Spectral fission can be achieved in a number of quite different ways in the frequency domain. *Spectral arpeggiation* is a process that draws our attention to the individual spectral components by isolating, or emphasising, each in sequence. This can be achieved purely vocally over a drone pitch by using appropriate vowel formants to emphasise partials above the fundamental. The computer can apply this process to any sound-source, even whilst it is in motion. (**Sound example 3.20**).

Spectral tracing strips away the spectral components in order of increasing loudness (**Appendix p25**). When only a few components are left, any sound is reduced to a delicate tracery of (shifting) sine-wave constituents. Complexly varying sources produce the most fascinating results as those partials which are at any moment in the permitted group (the loudest) change from window to window. We hear new partials entering (while others leave) producing "melodies" internal to the source sound. This feature can often be enhanced by *time-stretching* so that the rate of partial change is slowed down. *Spectral tracing* can also be done in a time-variable manner so that a sound gradually dissolves into its internal sine-wave tracery. (**Sound example 3.21**).

Spectral time-stretching, which we will deal with more fully in Chapter 11, can produce unexpected spectral consequences when applied to noisy sounds. In a noisy sound the spectrum is changing too quickly for us to gain any pitch or inharmonic multi-pitched percept from any particular time-window. Once, however, we slow down the rate of change the spectrum becomes stable or stable-in-motion for long enough for us to hear out the originally instantaneous window values. In general, these are inharmonic and hence we produce a "metallic" inharmonic (usually moving) ringing percept. By making perceptible what was not previously perceptible we effect a "magical" transformation of the sonic material. Again, this can be effected in a time-varying manner so that the inharmonicity emerges gradually from within the stretching sound. (**Sound example 3.22**).

Alternatively we may elaborate the spectrum in the time-domain by a process of constructive distortion. By searching for wavesets (zero-crossing pairs: **Appendix p50**) and then repeating the wavesets before proceeding to the next (*Waveset time-stretching*) we may time stretch the source without altering its pitch (see elsewhere for the limitations on this process). (**Appendix p55**). Wavesets correspond to wavecycles in many pitched sounds, but not always (**Appendix p50**). Their advantage in the context of constructive distortion is that very noisy sounds, having no pitch, have no true wavecycles – but we can still segment them into wavesets (**Appendix p50**).

In a very simple sound source (e.g. a steady waveform, from any oscillator) waveset time-stretching produces no artefacts. In a complexly evolving signal (especially a noisy one) each waveset will be different, often radically different, to the previous one, but we will not perceptually register the content of that waveset in its own right (see the discussion of time-frames in Chapter 1). It merely contributes to the more general percept of noisiness. The more we repeat each waveset however, the closer it comes to the grain threshold where we can hear out the implied pitch and the spectral quality implicit in

its waveform. With a 5 or 6 fold repetition therefore, the source sound begins to reveal a lightning fast stream of pitched beads, all of a slightly different spectral quality. A 32 fold repetition produces a clear "random melody" apparently quite divorced from the source. A three or four fold repetition produces a "phasing"-like aura around the sound in which a glimmer of the bead stream is beginning to be apparent. (**Sound example 3.23**).

Again, we have a compositional process which makes perceptible aspects of the signal which were not perceptible. But in this case, the aural result is entirely different. The new sounds are time-domain *artefacts* consistent with the original signal, rather than revelations of an intrinsic internal structure. For this reason I refer to these processes as constructive distortion.

SPECTRAL MANIPULATION IN THE FREQUENCY DOMAIN

There are many other processes of spectral manipulation we can apply to signals in the frequency domain. Most of these are only interesting if we apply them to moving spectra because they rely on the interaction of data in different (time) windows – and if these sets of data are very similar we will perceive no change.

We may select a window (or sets of windows) and freeze either the frequency data or the loudness data which we find there over the ensuing signal (*spectral freezing*). If the frequency data is held constant, the channel amplitudes (loudnesses) continue to vary as in the original signal but the channel frequencies do not change. If the amplitude data is held constant then the channel frequencies continue to vary as in the original signal. As mentioned previously, in a complex signal, holding the amplitude data is often more effective in achieving a sense of "freezing" the signal. We can also freeze both amplitude and frequency data but, with a complex signal, this tends to sound like a sudden *splice* between a moving signal and a synthetic drone. (**Sound example 3.24**).

We may average the spectral data in a each frequency-band channel over N time-windows (*spectral blurring*) thus reducing the amount of detail available for reconstructing the signal. This can be used to "wash out" the detail in a segmented signal and works especially effectively on spikey crackly signals (those with brief, bright peaks). We can do this and also reduce the number of partials (*spectral trace & blur*) and we may do either of these things in a time-variable manner so that the details of a sequence gradually become blurred or gradually emerge as distinct. (**Sound example 3.25**).

Finally, we may shuffle the time-window data in any way we choose (*spectral shuffling*), shuffling windows in groups of 8, 17, 61 etc. With large numbers of windows in a shuffled group we produce an audible rearrangements of signal segments, but with only a few windows we create another process of sound blurring akin to brassage, and particularly apparent in rapid sequences. (**Sound example 3.26**).

SPECTRAL MANIPULATION IN THE TIME-DOMAIN

A whole series of spectral transformations can be effected in the time-domain by operating on wavesets defined as pairs of zero-crossings. Bearing in mind that these do not necessarily correspond to true wavecycles, even in relatively simple signals, we will anticipate producing various unexpected artefacts in complex sounds. In general, the effects produced will not be entirely predictable, but they will be tied to the morphology (time changing characteristics) of the original sound. Hence the resulting sound will be clearly related to the source in a way which may be musically useful. As this process destroys the original form of the wave I will refer to it as *destructive distortion*. The following manipulations suggest themselves.

We may replace wavesets with a waveform of a different shape but the same amplitude (*waveset substitution* : **Appendix p52**). Thus we may convert all the wavesets to square waves, triangular waves, sine-waves, or even user-defined waveforms. Superficially, one might expect that sine-wave replacement would in some way simplify, or clarify, the spectrum. Again, this may be true with simple sound input but complex sounds are just changed in spectral "feel" as a rapidly changing sine-wave is no less perceptually chaotic than a rapidly changing arbitrary wave-shape. In the sound examples the sound with a wood-like attack has wavesets replaced by square waves, and then by sine waves. Two interpolating sequences (see Chapter 12) between the 'wood' and each of the transformed sounds is then created by *inbetweening* (see **Appendix p46** & Chapter 12). **(Sound example 3.27)**.

Inverting the half-wave-cycles (*waveset inversion*: **Appendix p51**) usually produces an "edge" to the spectral characteristics of the sound. We might also change the spectrum by applying a power factor to the waveform shape itself (*waveset distortion*: **Appendix p52**) **(Sound example 3.28)**.

We may average the waveset shape over N wavesets (*waveset averaging*). Although this process appears to be similar to the process of spectral blurring, it is in fact quite irrational, averaging the waveset length and the wave shape (and hence the resulting spectral contour) in perceptually unpredictable way. More interesting (though apparently less promising) we may replace N in every M wavesets by silence (*waveset omission* : **Appendix p51**). For example, every alternate waveset may be replaced by silence. Superficially, this would appear to be an unpromising approach but we are in fact thus changing the waveform. Again, this process introduces a slightly rasping "edge" to the sound quality of the source sound which increases as more "silence" is introduced. **(Sound example 3.29)**.

We may add 'harmonic components' to the waveset in any desired proportions (*waveset harmonic distortion*) by making copies of the waveset which are 1/2 as short (1/3 as short etc) and superimposing 2 (3) of these on the original waveform in any specified amplitude weighting. With an elementary waveset form this adds harmonics in a rational and predictable way. With a complex waveform, it enriches the spectrum in a not wholly predictable way, though we can fairly well predict how the spectral energy will be redistributed. **(Appendix p52)**.

We may also rearrange wavesets in any specified way (*waveset shuffling* : **Appendix p51**) or reverse wavesets or groups of N wavesets (*waveset reversal* : **Appendix p51**). Again, where N is large we produce a fairly predictable brassage of reverse segments, but with smaller values of N the signal is altered in subtle ways. Values of N at the threshold of grain perceptibility are especially interesting. Finally, we may introduce small, random changes to the waveset lengths in the signal (*waveset shaking* : **Appendix p51**). This has the effect of adding "roughness" to clearly pitched sounds.

Such distortion procedures work particularly well with short sounds having distinctive loudness trajectories. In the sound example a set of such sounds, suggesting a bouncing object, is destructively distorted in various ways, suggesting a change in the physical medium in which the 'bouncing' takes place (e.g. bouncing in sand). **(Sound example 3.30)**.

CONCLUSION

In a sense, almost any manipulation of a signal will alter its spectrum. Even *editing* (most obviously in very short time-frames in *brassage* e.g.) alters the time-varying nature of the spectrum. But, as we have already made clear, many of the areas discussed in the different chapters of this book overlap considerably. Here we have attempted to focus on sound composition in a particular way, through the concept of "spectrum". Spectral thinking is integral to all sound composition and should be borne in mind as we proceed to explore other aspects of this world.

CHAPTER 4

ONSET

WHAT IS SIGNIFICANT ABOUT THE ONSET OF A SOUND ?

In the previous Chapter (Spectrum) we have discussed properties of sounds which they possess at every moment, even though these properties may change from moment to moment. There are, however, properties of sound intrinsically tied to *the way in which the sound changes*. In this Chapter and the next we will look at those properties. In fact, the next chapter, entitled "Continuation" might seem to deal happily with all those properties. Why should we single out the properties of the onset of a sound, its attack, as being any different to those that follow?

The onset of a sound, however, has two particular properties which are perceptually interesting. In most naturally occurring sounds the onset gives us some clue as to the causality of the sound – what source is producing it, how much energy was expended in producing it, where it's coming from. Of course, we can pick up some of this information from later moments in the sound, but such information has a primitive and potentially life-threatening importance in the species development of hearing. After all, hearing did not develop to allow us to compose music, but to better help us to survive. We are therefore particularly sensitive to the qualities of sound onset – at some stage in the past our ancestors lives may have depended on the correct interpretation of that data. A moment's hesitation for reflection may have been too long!

Secondly, because of the way sound events are initiated in the physical world, the onset moment almost inevitably has some special properties. Thus a resonating cavity (like a pipe) may produce a sustained and stable sound once it is activated, but there needs to be a moment of transition from non-activation to activation, usually involving exceeding some energy threshold, to push the system into motion. Bells need to be struck, flutes blown etc. Some resonating systems can, with practice, be put into resonance with almost no discontinuity (the voice, bowl gongs, bloogles). Others either require a transient onset event, or can be initiated with such an event, (flute or brass tonguing). Other systems have internal resonance – once set in motion we do not have to continue supplying energy to them – but we therefore have to supply a relatively large amount of energy in a short time at the event onset (piano-string, bell). Other systems produce intrinsically short sounds as they have no internal resonance. Such sources can produce either individual short sounds (drums, xylophones, many vocal consonants) or be activated iteratively (drum roll, rolled "r", low contrabassoon notes).

Iterative sounds are a special case in which perceptual considerations enter into our judgement. Low and high contrabassoon notes are both produced by the discontinuous movement of the reed. However, in the lower notes we hear out those individual motions as they individually fall within the grain time-frame (See Chapter 1). Above a certain speed, the individual reed movements fall below the grain time-frame boundary and the units meld in perception into a continuous event. Sounds which are perceptually iterative, or granular, can be thought of as a sequence of onset events. This means that they have special properties which differentiate them (perceptually) from continuous sounds and must be treated differently when we compose with them. These matters are discussed in Chapters 6,7 and 8.

With acoustic instruments the initiating transition from "off" to "on" is most often a complex event in its own right, a clang, a breathy release, or whatever, with the dimensions of a grain (See Chapter 1) and with its own intrinsic sonic properties. These properties are, in fact, so important that we can destroy the recognisability of instrumental sounds (flute, trumpet, violin) fairly easily by removing their onset portion. It is of course not only the onset which is involved in source recognition. Sound sources with internal resonance and natural decay (struck piano strings, struck bells) are also partly recognisable through this decay process and if it is artificially prevented from occurring, our percept may change (is it a piano or is it a flute?). For a more detailed discussion see *On Sonic Art*.

We need, therefore, to pay special attention to the onset characteristics of sounds.

GRAIN-SCALE SOUNDS

Very short sounds (xylophone notes, vocal clicks, two pebbles struck together) may be regarded as onsets without continuation. Such sounds may be studied as a class on their own. We may be aware of pitch, pitch motion, spectral type (harmonicity, inharmonicity, noisiness etc) or spectral motion. But our percept will also be influenced strongly by the loudness trajectory of such sounds. (See **Diagram 1**).

Thus any grain-scale sound having a loudness trajectory of type 1a (see Diagram) will appear "struck" as the trajectory implies that all the energy is imparted in an initial shock and then dies away naturally. We can create the percept "struck object" by imposing such a brief loudness trajectory on almost any spectral structure. For example, a *time-stretched* vocal sound may have an overall trajectory imposed on it made out of such grain-scale trajectories, but repeated. The individual grains of the resulting iterated sound may appear like struck wood (**Sound example 4.1**).

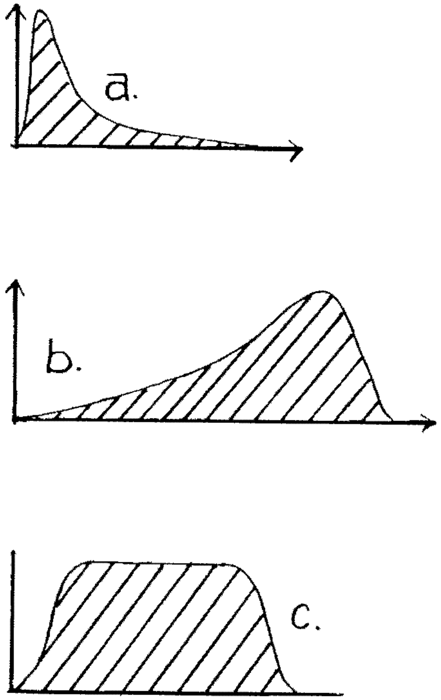
If these grains are then spectrally altered (using, for example, the various *destructive distortion* instruments discussed in Chapter 3) we may alter the perceived nature of the "material" being "struck". In particular, the more noisy the spectrum, the more "drum-like" or "cymbal-like" but we are retaining the percept "struck" because of the persisting form of the loudness trajectory. (**Sound example 4.2**).

If, however, we provide a different loudness trajectory (by *enveloping*) like type 1b, which has a quiet onset and peaks near the end, the energy in the sound seems to grow, which we might intuitively associate with rubbing or stroking or some other gentler way of coaxing an object into its natural vibrating mode. At the very least the percept is "gradual initiated", rather than "sudden initiated". (**Sound example 4.3**).

Yet another energy trajectory, a sudden excitation brought to an abrupt end (1c), suggests perhaps an extremely forced scraping together of materials where the evolution of the process is controlled by forces external to the natural vibrating properties of the material e.g. sawing or bowing, particularly where these produce forced vibrations rather than natural resonant frequencies. (**Sound example 4.4**).

Thus with such very brief sound, transformation between quite different aural percepts can be effected by the simple device of altering loudness trajectory. Combining this with the control of spectral content and spectral change gives us a very powerful purchase on the sound-composition of grain-size sounds.

DIAGRAM 1



ALTERED PHYSICALITY

What we have been describing are modifications to the perceived physicality of the sound. If we proceed now to sounds which also have a continuation, subtle alterations of the onset characteristics may still radically alter the perceived physicality of the sound. For example we can impose a sudden-initiated (struck-like) onset on any sound purely by providing an appropriate onset loudness trajectory, or we can make a sound gradual-initiated by giving it an onset loudness trajectory which rises more slowly. In the grain time-frame we can proceed from the "struck inelastic object" to the "struck resonating object" to the "rubbed" and beyond that to the situation where the sound appears to rise out of nowhere like the "singing" of bowl gongs or bloogles. (**Sound example 4.5**).

Where the sound has a "struck" quality, we may imply not just the energy but the physical quality of the striking medium. Harder striking agents tend to excite higher frequencies in the spectrum of the vibrated material (compare padded stick, rubber headed sticks and wooden sticks on a vibraphone). We can generalise this notion of physical "hardness" of an onset, to the onset of any sound. By exaggerating the high frequency components in the onset moment we create a more "hard" or "brittle" attack. (I make no apologies for using these qualitative or analogical terms. Grain time-frame events have an indivisible qualitative unity as percepts. We can give physical and mathematical correlates for many of the perceived properties, but in the immediate moment of perception we do not apprehend these physical and mathematical correlates. These are things we learn to appreciate on reflection and repeated listening.)

One way to achieve this attack hardening is to *mix* octave upward transposed copies of the source onto the onset moment, with loudness trajectories which have a very sudden onset and then die away relatively quickly behind the original sound (we do not have to use octave transposition and the rate of decay is clearly a matter of aesthetic intent and judgement). (*octave stacking*). The transpositions might be in the time frame of the original sound, or time-contracted (as with *tape-speed variation* : See Chapter 11). The latter will add new structure to the attack, particularly if the sound itself is quickly changing. We can also, of course, enhance the attack with downward transpositions of a sound, with similar loudness trajectories, the physical correlate of such a process being less clear. This latter fact is not necessarily important as, in sound composition, we are creating an artificial sonic world. (**Sound example 4.6**).

We can, for example, achieve in this way a "hard" or "metallic" attack to a sound which is revealed (immediately) in its continuation to be the sound of water, or human speech, or a non-representational spectrum suggesting physical softness and elasticity. We are not constrained by the photographically real, but our perception is guided by physical intuition even when listening to sound made in the entirely contrived space of sound composition. (**Sound example 4.7**).

Another procedure is to add noise to the sound onset but allow it to die away very rapidly. We may cause the noisiness to 'resolve' onto an actual wavelength of the modified sound by preceding that wavelength with repetitions of itself which are increasingly randomised i.e. noisier. This produces a plucked-string-like attack (*sound plucking*) and relates to a well known synthesis instrument for producing plucked string sounds called the Karplus Strong algorithm. (**Sound example 4.8**).

The effect of modifying the onset has to be taken into consideration when other processes are put into motion. In particular, *time-stretching* the onset of a sound will alter its loudness trajectory and may even extend it beyond the grain time-frame. As the onset is so perceptually significant,

time-stretching the onset is much more perceptually potent than time-stretching the continuation. This issue is discussed in Chapter 1. Also *editing* procedures on sequences (melodies, speech-stream etc.) in many circumstances need to preserve event onsets if they are not to radically alter the perceived nature of the materials (the latter, of course, may be desired). Finally, extremely dense textures in mono will eventually destroy onset characteristics, whereas stereo separation will allow the ear to discriminate event onsets even in very dense situations. (**Sound example 4.9**).

ALTERED CAUSALITY

Because the onset characteristics of a sound are such a significant clue to the sound's origin, we can alter the causality of a sound through various compositional devices. In particular, a sound with a vocal onset tends to retain its "voiceness" when continuation information contradicts our initial intuitive assumption. The piece *Vox-5* uses this causality transfer throughout in a very conscious way but it can operate on a more immediate level.

Listen first to **Sound example 4.10**. A vocally initiated event transforms in a strange (and vocally impossible) way. If we listen more carefully, we will hear that there is a *splice* (in fact a splice at a zero crossing: *zero-cutting*) in this sound where the vocal initiation is *spliced* onto its non-vocal (but voice derived) continuation. (In *Vox-5* the vocal/non-vocal transitions are achieved by smooth *spectral interpolation*, rather than abrupt *splicing* - see Chapter 12). When this abrupt change is pointed out to us, we begin to notice it as a rather obvious discontinuity, the "causal chain" is broken, but in the wider context of a musical piece, using many such voice initiated events, we may not so easily hone in on the discontinuity.

A more radical causality shift can be produced by onset fusion. When we hear two sounds at the same time certain properties of the aural stream allow us to differentiate them. Even when we hear two violinists playing in unison, we are aware that we are hearing two violins and not a single instrument producing the same sound stream. At least two important factors in our perception permit us to differentiate the two sources. Firstly, the micro fluctuations of the spectral components from one of the sources will be precisely in step with one another but generally out of step with those of the other source. So in the continuation we can aurally separate the sources. Secondly, the onset of the two events will be slightly out of synchronisation no matter how accurately they are played. Thus we can aurally separate the two sources in the onset moments.

If we now precisely align the onsets of two (or more) sounds to the nearest sample (*onset synchronisation*) our ability to separate the sources at onset is removed. The instantaneous percept is one of a single source. However, the continuation immediately reveals that we are mistaken. We thus produce a percept with "dual causality". At its outset it is one source but it rapidly unfolds into two.

In **Sound example 4.11** from *Vox-5* this process is applied to three vocal sources. Listen carefully to the first sound in the sequence. The percept is of "bell" but also voices, even though the sources are only untransformed voices. This initial sound initiates a sequence of similar sounds, but as the sequence proceeds the vocal sources are also gradually *spectrally stretched* (See Chapter 3) becoming more and more bell-like in the process.

CHAPTER 5

CONTINUATION

WHAT IS CONTINUATION ?

Apart from grain–duration sounds, once a sound has been initiated it must continue to evolve in some way. Only contrived synthetic sounds remain completely stable in every respect. In this chapter we will discuss various properties of this sound continuation, sometimes referred to as morphology and allure.

Some types of sound–continuation are, however, quite special. Sounds made from a sequence of perceived rapid onsets (grain–streams), sounds made from sequences of short and different elements (sequences) and sound which dynamically transform one set of characteristics into a quite different set (dynamic interpolation) all have special continuation properties which we will discuss in later chapters. Here we will deal with the way in which certain single properties, or a small set of properties, of a sound may evolve in a fairly prescribed way as the sound unfolds. These same properties may evolve similarly over grain–streams, sequences and dynamically interpolating sounds. They are not mutually exclusive.

DISPERSIVE CONTINUATION & ITS DEVELOPMENT

Certain natural sounds are initiated by a single or brief cause (striking, short blow or rub) and then continue to evolve because the physical material involved has some natural internal resonance, (stretched metal strings, bells) or the sound is enhanced by a cavity resonance (slot drum, resonant hall acoustics). As the medium is no longer being excited, however, the sound will usually gradually become quieter (not inevitably; for example the sound of the tam–tam may grow louder before eventually fading away) and its spectrum may gradually change. In particular, higher frequencies tend to die away more quickly than lower frequencies except in cases where a cavity offers a resonating mode to a particular pitch, or pitch area, within the sound. This may then persist for longer than any other pitch component not having such a resonating mode. We will call this mode of continuation *attack dispersal*. (**Sound example 5.1**).

In the studio we can immediately reverse this train of events (*sound reversal*), causing the sound to grow from nowhere, gradually accumulating all its spectral characteristics and ending abruptly at a point of (usually) maximum loudness. The only real–world comparable experience might be that of a sound–source approaching us from a great distance and suddenly stopping on reaching our location. We will call this type of continuation an *accumulation*.

Accumulations are more startling if they are made from non–linear dispersals. The decay of loudness and spectral energy of a piano note tends to be close to linear, so the associated accumulation is little more than a crescendo. Gongs or tam–tams or other complex–spectra events (e.g. the resonance of the undamped piano frame when struck by a heavy metal object) have a much more complex dispersal in which many of the initial high frequency components die away rapidly. The associated accumulation therefore begins very gradually but accelerates in spectral "momentum" towards the end, generating a growing sense of anticipation. (**Sound example 5.2**).

The structures of dispersal and accumulation can be combined to generate more interesting continuation structures. By *splicing* copies of one segment of a dispersal sound in a back-to-back fashion so that the reversed version exactly dovetails into its forward version, we can create an ebb and flow of spectral energy (see **Diagram 1**). By time-variably time-stretching (*time-warping*)the result, we can avoid a merely time-cyclic ebb and flow. More significantly, the closer we cut to the onset of the original sound, the more spectral momentum our accumulation will gather before releasing into the dispersal phase. Hence we can build up a set of related events with different degrees of musical intensity or tension as the accumulation approaches closer and closer to the onset point. (See **Diagram 2**). As the listener cannot tell how close any particular event will approach, the sense of spectral anticipation can be played with as an aspect of compositional structure.

This reminds me somewhat of the Japanese gourmet habit of eating a fish which is poisonous if the flesh too close to the liver is eaten. Some officionados, out of bravado, ask for the fish to be cut within a hair's breadth of the liver, sometime with fatal consequences. Sound composition is, fortunately, a little less dangerous. (**Sound example 5.3**).

Again, *time-warping*, spatial motion or different types of spectral reinforcement (see Chapter 3) can be used to develop this basic idea.

UNDULATING CONTINUATION AND ITS DEVELOPMENT

Certain time varying properties of sound evolve in an undulating fashion. The most obvious examples of these are , undulation of pitch (vibrato) and undulation of loudness (tremolando). Undulating continuation is related to physical activities like shaking and it is no accident that a wide, trill-like, vibrato is known as a "shake". These variations in vocal sounds involve, in some sense, the physical shaking of the diaphragm, larynx or throat or (in more extended vocal techniques) the rib cage, the head or the whole body (!). This may also be induced in elastic physical objects (like thin metal sheets, thin wooden boards etc) by physically shaking them. (**Sound example 5.4**).

In naturally occuring vibrato and tremolando, there is moment-to-moment instability of undulation speed (frequency) and undulation depth (pitch-excursion for vibrato, loudness fluctuation for tremolando) which is not immediately obvious until we create artificial vibrato or tremolando in which these features are completely regular. Completely regular speed, in particular, gives the undulation a cyclical, or rhythmic, quality drawing our attention to its rhythmicity. (**Sound example 5.5**).

Both speed and depth of vibrato or tremolo may have an overall trajectory (e.g. increasing speed, decreasing depth etc.). In many non-Western art music cultures, subtle control of vibrato speed and depth is an important aspect of performance practice. Even in Western popular music, gliding upwards onto an Hpitch as vibrato is added, is a common phenomena. (**Sound example 5.6**).

Vocal vibrato is in fact a complex phenomenon. Although the pitch (and therefore the partials) of the sound drift up and down in frequency, for a given vowel sound the spectral peaks (formants) remain where they are, or, in dipthongs, move independently. (See **Appendix p10** and **p66**). The pitch excursions of vibrato thus make it more likely that any particular partial in a sound will spend at least a little of its existence in the relatively amplified environment of a spectral peak. Hence vibrato can be used to add volume to the vocal sound. (See **Diagram 3**).

DIAGRAM 1

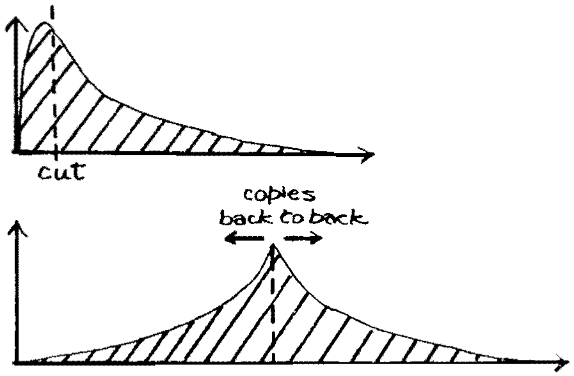
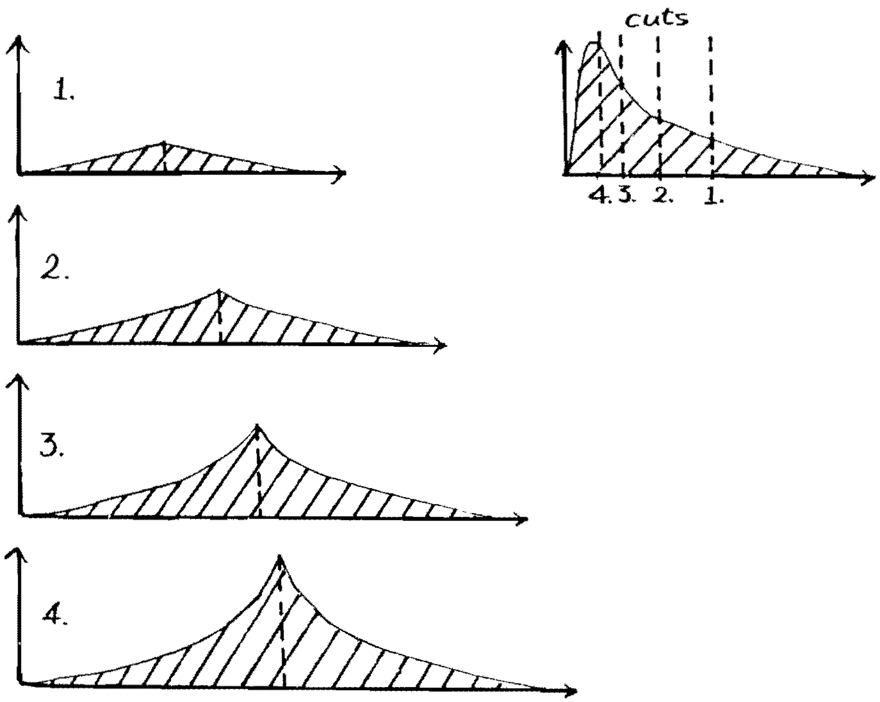


DIAGRAM 2



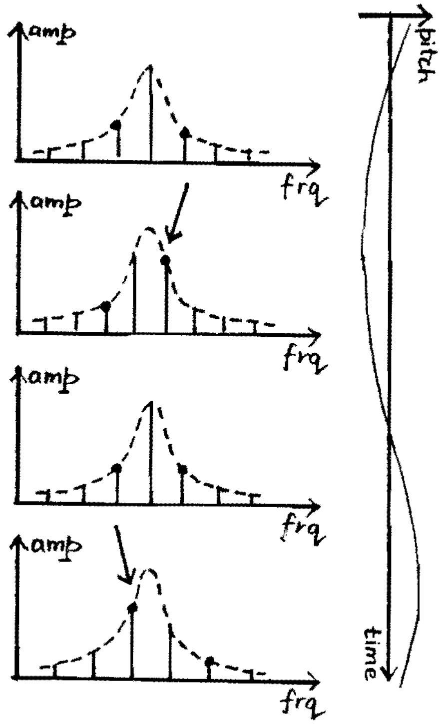


DIAGRAM 3

Ideally we should separate formant data before adding some new vibrato property to a sound, reimposing the separate motion of the formants on the pitch-varied sound. (See *formant preserving spectral manipulation* : **Appendix p17**). In practice we can often get away with mere *tape-speed variation* transposition of the original sound source.

We may extract the ongoing evolution of undulating properties using *pitch-tracking* (**Appendix pp70-71**) or *envelope following* (**Appendix p58**) and apply the extracted data to other events. We may also modify (e.g. *expanding, compression* : **Appendix p60**) the undulating properties of the source-sound. Alternatively, we may define undulations in pitch (*vibrato*) or loudness (*tremolo*) in a sound, with control over time-varying speed and depth. In this way we may, e.g. impose voice-like articulations on sounds with static spectra, or non-vocal environmental sources (e.g.the sound of a power drill). (**Sound example 5.7**).

We may also produce extreme, or non-naturally-occuring, articulations using extremely wide (several octaves) vibrato, or push tremolando to the point of sound granulation. At all stages the overall trajectories (gradual variation of speed and depth) will be important compositional parameters. (**Sound example 5.8**).

In the limit (very wide and slow) tremolando becomes large-scale loudness trajectory and we may progress from undulating continuation to forced articulation (and vice versa). Similarly, vibrato (very wide and slow) becomes the forced articulation of moving pitch. We thus move from the undulatory articulation of a stable Hpitch to the realm of pitch motion structures in which Hpitch has no further significance. (**Sound example 5.9**).

At the opposite extreme, tremolando may become so deep and fast, that the sound granulates. We may then develop a dense grain texture (see Chapter 8) on which we may impose a new tremolando agitation and this may all happen within the ongoing flow of a single event. (**Sound example 5.10**).

We may also imagine undulatory continuation of a sound's spectrum, fluctuating in its harmonicity\inharmonicity dimension, its stableness-noise dimension, or in its formant-position dimension (*spectral undulation*). Formantal undulations (like hand-yodelling, head-shake flutters or "yuyuyuyu" articulation) can be produced and controlled entirely vocally. (**Sound example 5.11**).

FORCED CONTINUATION AND ITS DEVELOPMENT

In any system where the energising source has to be continually applied to sustain the sound (bowed string, blown reed, speech) the activator exercises continuous control over the evolution of the sound. With an instrument, a player can force a particular type of continuation, a crescendo, a vibrato or tremolando, or an overblowing, usually in a time-controllable fashion. In general, the way in which the sound changes in loudness and spectrum (with scraped or bowed sounds etc) or sometimes in pitch (with rotor generation as in a siren or wind machine) will tell us how much energy is being applied to the sounding system. These forced loudness, spectral or pitch movement shapes may thus be thought of as physical gestures translated into sound.

Any gestural shape which can be applied by breath control or hand pressure on a wind or bowed string instrument, can be reproduced over any arbitrary sound (a sustained piano tone, the sound of a dense crowd) by applying an appropriate loudness trajectory (*enveloping*) with perhaps more subtle paralleling features (spectral enhancement by *filtering*, harmonicity shifting by *spectral stretching* or subtle *delay*). Moreover, these sonically created shapes can transcend the boundaries of the physically likely (!). Sound which are necessarily quiet in the real world (unvoiced whispering) can be unnervingly loud, while sounds we associate with great forcefulness, e.g. the crashing together of large, heavy objects, the forced grating of metal surfaces, can be given a pianissimo delicacy.

Moreover, we can extract the properties of existing gestures and modify them in musically appropriate ways. This is most easily done with time-varying loudness information which we can capture (*envelope following*) and modify using a loudness trajectory manipulation instrument (*envelope transformation*), reapplying it to the original sound (*envelope substitution*), or transferring it to other sounds (*enveloping* or *envelope substitution* : see **Appendix p59**).

Information can be extracted from instrumental performance (which we might specifically compose or improvise for the purpose), speech or vocal improvisation but also from fleeting unpredictable phenomena (the dripping of a tap) or working in stereo, the character of a whole field of naturally occuring events (e.g. traffic flow, the swarming of bats etc). The extracted gestural information can then be modified and reapplied to the same material (*envelope transformation* followed by *envelope substitution*), or applied to some entirely different musical phenomenon, (*enveloping* or *envelope substitution* : see **Appendix p59**) or stored for use at some later date. All such manipulations of the loudness trajectory are discussed more fully in Chapter 10.

Sounds may also have a specific spatial continuation. A whole chapter of *On Sonic Art* is devoted to the exploration of spatial possibilities in sound composition. Here we will only note that spatial movement can be a type of forced continuation applied to a sound, that it will work more effectively with some sounds than with others and that it can be transferred from one sound to another provided these limitations are borne in mind. Noisy sounds, grain-streams and fast sequences move particularly well, low continuous sounds particularly poorly. Some sounds move so well that rapid spatial movement (e.g. a rapid left right sweep) may appear as an almost indecipherable quality of grain or of sound onset. **(Sound example 5.12)**.

The movement from mono into stereo may also play a significant role in *dynamic interpolation* (See Chapter 12).

CONSTRUCTED CONTINUATION : TIME-STRETCHING & REVERBERATION

In many cases we may be faced with a relatively short sound and wish to create a continuation for it. There are a number of ways in which sounds can be extended artificially, some of which reveal continuation properties intrinsic to the sound (*reverberation*, *time-stretching*, some kinds of *brassage*) while others impose their own continuation properties (*zigzagging*, granular-extension and other types of brassage).

The most obvious way to create continuation is through time-stretching. There are several ways to do this (*brassage/harmoniser*, *spectral stretching*, *waveset time stretching*, *granular time stretching*) and these are discussed more fully in Chapter 11. It is clear, however, that through time-stretching we can expand the indivisible qualitative properties of a grain into a perceptibly time-varying structure – a continuation. In this way an indivisible unity can be made to reveal a surprising morphology. This

process also can change a rapid continuation into something in a phrase time-frame, for example, formant-gliding consonants ("w", "y" in English) become slow formant transformations ("oo" —> "uh" and "i" —> "uh"). Conversely, a continuation structure can be locked into the indivisible qualitative character of a grain, through extreme time-contraction. **(Sound example 5.13)**.

Continuation may also be generated through *reverberation*. This principle is used in many acoustic instruments where a sound box allows the energy from an instantaneous sound-event to be reflected around a space and hence sustained. Reverberation will extend any momentarily *stable* pitch and spectral properties in a short event. It may also be combined with *filtering* to sustain specific pitches or spectral bands. It provides a new dispersal continuation for the sound. **(Sound example 5.14)**.

Natural reverberation is heard when (sufficiently loud) sounds are played in rooms or enclosures of any kind, except where the reverberation has been designed out, as in an anechoic chamber. Natural reverberation is particularly noticeable in old stone buildings (e.g. churches) or tiled spaces like bathrooms or swimming pool enclosures. Reverberation in fact results from various delayed (due to travelling in indirect paths to the ear, by bouncing off walls or other surfaces) and spectrally altered (due to the reflection process taking place on different physical types of surface) versions of the sound being mixed with the direct sound as it reaches the ear. Such processes can be electronically mimicked. The electronic mimicry has the added advantage that we can specify the dimensions of unlikely or impossible spaces (e.g. infinite reverberation from an 'infinitely long' hall, the inside of a biscuit tin for an orchestra etc.) There is thus an enormous variety of possibilities for creating sound continuation through reverberation.

Reverberation can be used to add the ambience of a particular space to any kind of sound. In this case we are playing with the illusion of the physical nature of the acoustic space itself, the generalised continuation properties of our whole sound set. But it can also be used in a specific way to enlarge or alter the nature of the individual sound events, extending (elements of) the spectrum in time.

Reverberation cannot, however, by itself, extend any undulatory or forced continuation properties of a sound. On the contrary it will blur these. Moving pitch will be extended as a pitch band, loudness variations will simply be averaged in the new dispersal. We can, of course, post-hoc, add undulatory or forced continuation properties (*vibrato*, *tremolo*, *enveloping*) to a reverberation extended sound. These will in fact help to unify the percept initiator-reverberator as a single sound-event. The reverberation part of the sound will appear to be part of the sound production itself, rather than an aspect of a sound box or characteristic of a room. **(Sound example 5.15)**.

Initiator-reverberator models have in fact been used recently to build synthesis models of acoustic instruments from vibraphones, where the separation might seem natural, to Tubas.

CONSTRUCTED CONTINUATION : ZIGZAGGING & BRASSAGE

Continuation can be generated by the process of *zigzagging*. Here, the sound is read in a forward-backward-forward etc sequence, each reversal-read starting at the point where the preceding read ended. The reversal points may be specified to be anywhere in the sound. Provided we start the whole process at the sound's beginning and end it at its end, the sound will appear to be artificially extended. In the simplest case, the sound may oscillate between two fixed times in the middle of the sound until told to proceed to the end. This "alternate looping" technique was used in early commercial samplers to allow a short sampled sound to be sustained. It may work adequately where the chosen portion of the sound has a stable pitch and spectrum. In general, however, any segment of a sound, beyond grain time-frame, will have a continuation structure and looping over it will cause this structure to be heard as a mechanical repetition in the sound. (see **Appendix p43**). (**Sound example 5.16**).

Zigzagging, however, can move from any point to any other within the sound, constantly changing length and average position as it does so. We can therefore use the process to select spectral, pitch or loudness movements or discontinuities in a sound and focus attention on them by alternating repetition. By altering the zigzag points subtly from zig to zig, we may vary the length (and hence duration) over the repeated segment. In this way zigzagging can be used to generate non-mechanical undulatory properties, or (on a longer timescale) dispersal-accumulation (see above) effects, within a sound-continuation. (**Sound example 5.17**).

Sounds can also be artificially continued by *brassage*. In the brassage process successive segments are *cut* from a sound and then *respliced* together to produce a new sound. Clearly, if the segments are replaced exactly as they were cut, we will reproduce the original sound. We can extend the duration of the sound by cutting overlapping segments from the source, but not overlapping them in the goal sound. (See **Appendix p44–B**). (see note at end of chapter).

As discussed previously, grain time-frame segments will produce a simple time-stretching of the sound (*harmoniser* effect). Slightly larger segments may introduce a granular percept into the sound as the perceived evolving shapes of the segments are heard as repeated. (**Sound example 5.18**). The segment granulation of an already granular source may produce unexpected *phasing* or *delay* effects within the sound. Longer segments, especially when operating on a segmented source (a melody, a speech stream) will result in a systematic collage of its elements. With regular segment-size our attention will be drawn to the segment length and the percept will probably be repetitively rhythmic. However, we may vary the segment size, either progressively or at random, producing a less rhythmicized, collage type extension. (**Sound example 5.19**).

This idea of brassage can be generalised. Using non-regular grain size near to the grain time-frame boundary, the instantaneous articulations of the sound will be echoed in an irregular fashion adding a spectral (very short time-frame) or articulatory (brief but longer time-frame) aura to the time-stretched source. We may also permit the process to select segments from within a time-range measured backwards from the current position in the source-sound (see **Appendix p44–C**). In this way, echo percepts are randomised further. Subtly controlling this and the previous factors, we can extract rich fields of possibilities from the small features of sounds with evolving spectra (especially sequences, which present us with constantly and perceptibly evolving spectra). (**Sound example 5.20**).

Ultimately, we can make the range include the entire span of the sound up to the current position (see **Appendix p44–D**). Now, as we proceed, all the previous properties of the sound become grist to the mill of continuation-production. In the case e.g. of a long, melodic phrase which we brassage in this way using a relatively large segment-size (including several notes) we will create a new melodic stream including more and more of the notes in the original melody. The original melody will thus control the evolving HArmonic field of Hpitch possibilities in the goal sound. (**Sound example 5.21**). On a smaller time-frame, the qualities of a highly characteristic onset event can be redistributed over the entire ensuing continuation. (**Sound example 5.22**).

The brassage components may also be varied in pitch. If this is done at random over a very small range, the effect will be to broaden the pitch band or the spectral energy of the original sound. (**Sound example 5.23**). We may also cycle round a set of pitches in a very small band providing a subtle "stepped vibrato" inside the continuation (particularly if grain-size is slightly random-varied to avoid rhythmic constructs in perception). (**Sound example 5.24**). The pitch band can also be progressively broadened. (**Sound example 5.25**). The loudness of segments can also be varied in a random, progressive or cyclical way. We might also spatialise, or progressively spatialise, the segmented components, moving from a point source to a spread source. (**Sound example 5.26**).

Eventually, such evolved manipulations (and their combinations) force a continuous or coherently segmented source to dissociate into a mass of atomic events. The ultimate process of this type is *Sound shredding* which completely deconstructs the original sound and will be discussed in Chapter 7.

CONSTRUCTED CONTINUATION : GRANULAR RECONSTRUCTION

We may generalise the brassage concept further, taking us into the realm of *granular reconstruction*. As with brassage, our source sound is *cut* into segments which we may vary in duration, pitch or loudness. However, instead of merely resplicing these tail-to-tail, they are used as the elements of an evolving texture in which we can control the density and the time-randomisation of the elements. (See **Appendix p73**). (**Sound example 5.27**).

In this way we can overlay segments in the resulting stream, or, if segments are very short, introduce momentary silences between the grains. This process, especially where used with very tiny grains, is also known as *granular synthesis* (the boundaries between sound processing and synthesis are fluid) and we may expect the spectral properties and the onset characteristics of the grains to influence the quality of the resulting sound stream, alongside imposed stream characteristics (density, pitch spread, loudness spread, spatial spread etc.). This process then passes over into texture control, and is discussed more fully in Chapter 8. (**Sound example 5.28**).

With granular reconstruction, if we keep the range and pitch-spread small, we may expect to generate a time-stretched goal-sound which is spectrally thickened (and hence often more noisy, or at least less focussed), the degree of thickening being controlled by our choice of both density and pitch bandwidth. But as range, density and bandwidth are increased and segment duration varied, perhaps progressively, the nature of the source will come to have a less significant influence on the goal sound. It will become part of the overall field-properties of a texture stream (see Chapter 8). (**Sound example 5.29**).

Final note : in order to simplify the discussion in this chapter (and also in the diagrammatic appendix) Brassage has been described here as a process in which the cut segments are not overlapped (apart from the length of the edits themselves) when they are reassembled. In practice, normal brassage/harmoniser processes use a certain degree of overlap of the adjacent segments to ensure sonic continuity and avoid granular artefacts (where these are not required) in the resulting sound.

CHAPTER 6

GRAIN–STREAMS

DISCONTINUOUS SPECTRA

In this chapter and the next we will discuss the particular properties of sounds with perceptibly discontinuous spectra. The spectra of many sounds are discontinuous on such a small time– frame that we perceive the result as noise (or in fact as pitch if the discontinuities recur cyclically), rather than as a sequence of changing but definite sound events. Once, however, the individual spectral moments are stable or stable–in–motion for a grain time–frame or more, we perceive a sound–event with definite but rapidly discontinuous properties.

In one sense, all our sound–experience is discontinuous. No sound persists forever and, in the real world, will be interrupted by another, congruously or incongruously. We are here concerned with perceived discontinuous changes in the time range speed–of–normal–speech down to the lower limit of grain perception.

Compositionally, we tend to demand different things of discontinuous sounds, than of continuous ones. In particular, if we *time–stretch* a continuous sound, we may be disturbed by the onset distortion but the remainder of the sound may appear spectrally satisfactory. If we time–stretch a discontinuous sound, however, we will be disconcerted everywhere by onset distortion as the sound is a sequence of onsets. Often we want the sound (e.g. in the real environment, a drum roll, a speech–stream) to be delivered more slowly to us without the individual attacks (the drum strike, the timbral structure of consonants) being smeared out and hence transformed. We wish to be selective about what we time–stretch!

The idea of slowing down an event stream without slowing down the internal workings of the events is quite normal in traditional musical practice – we just play in a slower tempo on the same instrument – the internal tempi of the onset events is not affected. But with recorded sounds we have to make special arrangements to get this to work.

We will divide discontinuous sounds into two classes for the ensuing discussion. A grain–stream is a sound made by a succession of similar grain events. In the limit it is simply a single grain rapidly repeated. Even where this (non!) ideal limit is approached in naturally occuring sounds (e.g. low contrabassoon notes) we will discover that the individual grains are far from identical, nor are they ever completely regularly spaced in time. (**Sound example 6.1**).

Discontinuous sounds consisting of different individual units (speech, a melody on a single instrument, any rapid sequence of different events) we will refer to as sequences and will discuss these in the next chapter.

Both grain–streams and sequences can have (or can be composed to have) overall continuation properties (dispersive, undulatory and forced continuation and their developments, as discussed in Chapter 5). (**Sound example 6.2**). In this chapter and the next, we will talk only about those properties which are special to grain–streams and sequences.

CONSTRUCTING GRAIN STREAMS

Grain-streams appear naturally from iterative sound-production - any kind of roll or trill on drums, keyed percussion or any sounding material. They are produced vocally by rolled "r" sounds of various sorts in various languages, by lip-farts and by tracheal rasps. Vocally, such sounds may be used to modulate others (sung rolled "r", whistled rolled "r", flutter-tongued woodwind and brass etc). The rapid opening and closing of a resonating cavity containing a sound source (e.g. the hand over the mouth as we sing or whistle) can be used to naturally grain-stream any sound. **(Sound example 6.3)**.

In the studio, any continuous sound may be grain-streamed by imposing an appropriate on-off type loudness trajectory (*enveloping*), which itself might be obtained by *envelope following* another sound (see Chapter 10). (On-offness might be anything from a deep tremolando fluttering to an abrupt on-off *gating* of the signal). A particularly elegant way to achieve the effect is to generate a loudness trajectory on a sound tied to the wavesets or wavecycles it contains (*waveset enveloping*). If each on-off type trajectory is between about 25 and 100 wavesets in length, we will hear grain-streaming. (Below this limit, we may produce a rasping or spectral "coarsening" of the source sound). This process ties the grain-streaming to the internal properties of the source-sound so, for example, the grain- streaming ritardandos if the perceived pitch falls. This suggests to the ear that the falling portamento and the ritardando are causally linked and intrinsic to the goal-sound, rather than a compositional affectation (!). **(Sound example 6.4)**.

Alternatively grain-streams may be constructed by *splicing* together individual grains (!). *Looping* can be used to do this but will produce a mechanically regular result. An instrument which introduces random fluctuations of repetition-rate and randomly varies the pitch and loudness of successive grains over a small range (*iteration*) produces a more convincingly natural result **(Sound example 6.5)**.

More compositionally flexible, but more pernickety, is to use a *mixing* program so that individual grains can be placed and ordered, then repositioned, replaced or reordered using meta-instruments which allow us to manipulate mixing instructions (*mixshuffling*) or to generate and manipulate time-sequences (*sequence generation*).

In this way grain-streams can be given gentle accelerandos or ritardandos of different shapes and be slightly randomised to avoid a mechanistic result. **(Sound example 6.6)**. Similarly, the grains themselves can be sequentially modified using some kind of *enveloping* (see Chapter 4), or spectral transformation tools (e.g. destructive distortion through *waveset distortion*, *waveset harmonic distortion* or *waveset-averaging* : see Chapter 3) combined perhaps with *inbetweening* (see Chapter 12) to generate a set of intermediate sound-states. **(Sound example 6.7)**.

Short continuous sounds can be extended into longer grain-streamed sound by using *brassage* with appropriate (grain time-frame duration) segment-size. **(Sound example 6.8)**. The granulation of the resulting sound can be exaggerated by *corrugation* (see Chapter 10) and the regularity of the result mitigated by using some of the grain-stream manipulation tools to be described below.

Many of these compositional processes provide means of establishing audible (musical) links between materials of different types. We are able to link grains with grain-streams and continuous sounds with grain-streams in this way and hence begin to build networks of musical relationships amongst diverse materials.

DISSOLVING GRAIN-STREAMS

Just as continuous sounds can be made discontinuous, grain-streams can be dissolved into continuous, or even single-grain, forms. Speeding up a grain-stream by *tape speed variation* or *spectral time-contraction* (with no grain pitch alteration) may force the grain separation under the minimum time limit for grain-perception and the granulation frequency will eventually emerge as a pitch. **(Sound example 6.9)**. Alternatively, by speeding up the sequence rate of grains without changing the grains (*granular time-shrinking* : see below) we will breach the grain-perception limit. The sound will gradually become a continuous fuzz. In this case a related pitch may or may not emerge. **(Sound example 6.10)**. *Reverberation* will blur the distinction between grains. **(Sound example 6.11)**. Increasing the grain density (e.g. via the parameters of a *granular synthesis* instrument) will also gradually fog over the granular property of a grain-stream. **(Sound example 6.12)**.

Grain-streams may also be dissociated in other ways....

(1) slowing down the sequence of grains but not the grains themselves so that grains become detached events in their own right (*granular time-stretching* : **Sound example 6.13**).

(2) slowing down the sequence *and* the grains, so the internal morphology of the individual events comes to the foreground of perception (*spectral time-stretching* : **Sound example 6.14**).

(3) gradually shifting the pitches or spectral quality of different grains differently, so the grain-stream becomes a sequence (*granular reordering* : **Sound example 6.15**).

Again, we are describing here ways in which networks of musical relationships can be established amongst diverse musical materials.

CHANGING GRAIN-STREAM STRUCTURE

In some ways, a grain-stream is akin to a note-sequence on an instrument. In the latter we have control over the timing and Hpitching and sequencing of the events. In a grain-stream not constructed from individually chosen grains, but e.g. by enveloping a continuous source, we do not initially have this control. However, we would like to be able to treat grains in a similar way to the way we deal with note events. By the appropriate use of *gating*, *cutting* and *resplicing* or *remixing*, which may be all combined in a single sound processing instrument, we can retime the grains in a sequence using various mathematical templates (slow by a fixed factor, ritardando arithmetically, geometrically or exponentially, randomise grain locations about their mean, shrink the time-frame in similar ways and so on : *granular time-warping*). The grain-stream can thus be elegantly time-distorted without distorting the grain-constituents. **(Sound example 6.16)**.

We can also reverse the grain order (*granular reversing*), without reversing the individual grains themselves (*sound reversing*), producing what a traditional composer would recognise as a retrograde (as opposed to an accumulation : see Chapter 5). Thus a grain-stream moving upwards in pitch would become a grain-stream moving downwards in pitch. **(Sound example 6.17)**.

We can go further than this. We might rearrange the grains in some new order (*grain reordering*). A rising succession of pitched sounds might thus be converted into a (motivic) sequence. (**Sound example 6.18**). Or we might move the pitch of individual grains without altering the time-sequence of the stream, or alter the timing on individual grains without altering their pitch, or some combination of both. In this way the control of grain-streams passes over into more conventional notions of event sequencing, melody and rhythm. I will not dwell on these here because hundreds of books have already been written about melody and rhythm, whereas sound composition is a relatively new field. We will therefore concentrate on what can be newly achieved, assuming that what is already known about traditional musical parameters is already part of our compositional tool kit.

CHAPTER 7

SEQUENCES

MELODY & SPEECH

The most common examples of sequences in the natural world are human speech, and melodies played on acoustic instruments. However any rapidly articulated sound stream can be regarded as a sequence (some kinds of birdsong, klangfarbenmelodie passing between the instruments of an ensemble, a "break" on a multi instrument percussion set etc). We can also construct disjunct sequences of arbitrary sounds by simply *splicing* them together (e.g. a dripping tap, a car horn, an oboe note, a cough – with environmentally appropriate *or inappropriate* loudness and/or *reverberation* relations one to the other) or by modifying existing natural sequences (time-contraction of speech or music, for example). (**Sound example 7.1**).

Naturally occuring sequences cannot necessarily be accurately reproduced by *splicing* together (in the studio) constituent elements. The speech stream in particular has complex transition properties at the interfaces between different phonemes which are (1994) currently the subject of intensive investigation by researchers in speech synthesis. To synthesize the speech stream it may be more appropriate to model all the transitions between the elements we tend to notate in our writing systems, rather than those elements themselves (this is Diphone synthesis). Starting from the separate elements themselves, to achieve the flowing unity of such natural percepts as speech, it may be necessary to "massage" a purely spliced-together sequence. A simple approach might be to add a little subtle *reverberation*. However, for the present discussion, we will ignore this subtle flow property and treat all sequences as if they were formally equivalent.

Clearly, sequences of notes on a specific instrument and sequences of phonemes in a natural language have well-documented properties, but here we would like to consider the properties of any sequence whatsoever.

CONSTRUCTING & DESTRUCTING SEQUENCES

Sequences can be generated in many ways, apart from *splicing* together the individual elements "by hand". Any sound source in directed motion (e.g. a pitch-glide or a formant glide) can be *spliced* into elements which, when rearranged, do not retain the spectral-continuity of the original. (**Sound example 7.2**). An existing speech-stream can be similarly reordered to destroy the syntactic content and (depending on where we *cut*) the phoneme-continuity. We may do this by chopping up the sequence into conjunct segments and reordering them (as in *sound shredding* : see below), or by selecting segments to *cut*, at random (so they might overlap other chosen segments) and reordering them as they are *spliced* back together again (*random cutting* : **Appendix p41**). (**Sound example 7.3**).

We may work with a definite segment length or with arbitrary lengths and we may shift the loudness or pitch of the materials, marginally or radically, before constructing the new sequence. Alternatively we may *cut* our material into sequential segments, modify each in a non-progressive manner (different *filterings*, pitch shifts, etc) and reconstitute the original sequence by *resplicing* the elements together again, but they will now have discontinuously varying imposed properties. (**Sound example 7.4**).

Using *brassage* techniques, we may achieve similar results, if the segment size is set suitably large and we work on a clearly spectrally-evolving source. Using several brassages with similar segment length settings but different ranges (see **Appendix p44-C**) we might create a group of sequences with identical field properties but different, yet related, order properties i.e. smaller ranges would tend to preserve the order relations of the source, large ranges would reveal the material in the order sequence of the source but, in the meantime, return unpredictably to already revealed materials. (**Sound example 7.5**).

Brassage with pitch variation of segments over a finite Hpitch set (possibly cyclically) could establish an Hpitch-sequence from a pitch-continuous (or other) source. (**Sound example 7.6**). Even brassage with spatialisation (see **Appendix p45**) will separate a continuous (or any other) source into a spatial sequence and so long as they are clearly delineated, such spatial sequences can be musically manipulated (**Sound example 7.7**). A sequence from left to right can become a sequence from right to left, or a sequence generally moving to the left from the right but with deviations, can be restationed at a point in space, or change to an alternation of spatial position and so on.

The perception of sequence can also be destroyed in various ways. Increasing the speed of a sequence beyond a certain limit will produce a gritty noise or even, in the special circumstance of a sequence of regularly spaced events with strong attacks, a pitch percept. Conversely, *time-stretching* the sequence beyond a certain limit will bring the internal properties of the sequence into the phrase time-frame and the sequence of events will become a formal property of the larger time-frame. (**Sound example 7.8**).

Alternatively, copies of the sequence, or (groups of) its constituents may be amassed in textures (see Chapter 8) of sufficient density that our perception of sequence is overwhelmed. (**Sound example 7.9**).

Conversely a sequence may be shredded (*sound shredding*). In this process the sequence is *cut* into random-length conjunct segments which are then reordered randomly. This process may be repeated ad infinitum, gradually dissolving all but the most persistent spectral properties of the sequence in a watery complexity (the complex in fact becomes a simple unitary percept). (See **Appendix p41**). (**Sound example 7.10**).

GENERAL PROPERTIES OF SEQUENCES

All sequences have two general properties. They can define both a *field* and an *order*. Thus, on the large scale, the set of utterances with a particular natural language defines a field, the set of phonemes which may be used in the language. Similarly any sequence played on a piano defines the tuning set or HArmonic field of the piano (possibly just a subset of it). It is possible to construct sequences which do not have this property (in which no elements are repeated) just as it is possible to construct pitch fields where no Hpitch reference-frame is set up. But in general, for a finite piece of music, we will be working within some field, a reference-frame for the sequence constituents. (**Sound example 7.11**).

Sequences are also characterised by order properties. In existing musical languages, certain sequences of notes will be commonplace, others exceedingly unlikely. In a particular natural language certain clusterings of consonants (e.g. "scr" in English) will be commonplace, others rare and yet others absent. It is easy to imagine and to generate unordered sequences, though the human mind's pattern seeking predilection makes it over-deterministic, hearing definite pattern where pattern is only hinted at!

In the finite space of a musical composition we may expect a reference set (field) and ordering properties to be established quite quickly if they are to be used as musically composed elements. On the larger scale, these may be predetermined by cultural norms, like tuning systems or the phoneme set of a natural language, but traditional musical practice is usually concerned with working on subsets of these cultural norms and exploring the particular properties and malleabilities of these subsets.

In this context the size of the field is significant. Musical settings of prose, for example, may treat the Hpitch (and duration) material in terms of a small ordered reference set which is constantly regrouped in subsets (e.g. chord formations over a scale) and reordered (motivic variation), whereas the phonetic material establishes no such small time-frame field and order properties – the text is used as referential language, and field and order properties are on the very large timescale of extensive language utterance. In this situation, the text is perceived as being in a separate domain to the "musical".

Poetry, however, through assonance, alliteration and particularly rhyme, begins to adopt the small-scale reference-frame and order sequencing for phonemes we find normal in traditional musical practice. We therefore discover a meeting ground between phonemic and traditional Hpitch and durational musical concerns and these connections have been explored by sound poets (Amirkhanian etc) and composers (Berio etc) alike. As we move towards poetry which is more strongly focused in the sonority of words, or just of syllabic utterance, the importance of small-scale reference-frame and order sequencing may become overriding (e.g. Schwitters *Ursonata*). **(Sound example 7.12)**.

COMPOSING FIELD PROPERTIES

Constructing sequences from existing non-sequenced, or differently sequenced, objects (a flute melody, or an upward sweeping noise-band, or a traffic recording in a tunnel with a particular strong resonance, or a conversation in Japanese) ensures that some field properties of the source sounds will inhere in the resulting sequence: a defined Hpitch reference set and a flute spectrum (with or without onset characteristics; in the latter case the field is altered), rising noise-bands within a given range, the resonance characteristics of the tunnel, the spectral characteristics of the phonemes of Japanese and perhaps the sex characteristics of the specific voices. These field properties may then define the boundaries of the compositional domain (a piano is a piano is a piano is a piano) or conversely become part of the substance of it, as we transform the field characteristics (piano –> "bell" –> "gong" –> "cymbal" –> unvoiced sibilant etc.). **(Sound example 7.13)**.

Compositionally we can transform the field (reference-set) of a sequence through time by gradual substitution of one element for another or by the addition of new elements or reduction in the total number of elements (this can be done with or without a studio!). We can also gradually transform each element (e.g. by *destructive distortion* with *inbetweening*, see Chapter 12) so that the elements of a sequence become more differentiated or, conversely, more and more similar moving towards a grain-stream (see above), or simply different. **(Sound example 7.14)**.

Or we may blur the boundaries between the elements through processes like *reverberation*, *delay*, small time-frame *brassage*, *spectral blurring*, *spectral shuffling*, *waveset shuffling*, or *granular reconstruction*), or simply through *time-contraction* (of various sorts) so that the sequence succession rate falls below the grain time-frame and the percept becomes continuous. We thus move a sequence

towards a simple continuation sound. In these ways we can form bridges between sequences, grain streams and smoothly continuous sounds, establishing audible links between musically diverse materials. **(Sound example 7.15)**.

We may also *focus* the field-properties of a sequence, or part of a sequence by *looping* (repeating that portion). Thus a series of pitches may seen arbitrary (and may in fact be entirely random across the continuum, establishing no reference frame of Hpitch). If, however, we repeat any small subset of the sequence, it will very rapidly establish itself as its own reference set. We can thus move very rapidly from a sense of Hpitch absence to Hpitch definiteness. **(Sound example 7.16)**.

The same set-focusing phenomena applies to the spectral characteristics of sounds. If a meaningful group of words is repeated over and over again we eventually lose contact with meaning in the phrase as our attention focuses on the material as just a sequence of sonic events defining a sonic reference frame. **(Sound example 7.17)**.

COMPOSING ORDER PROPERTIES

We can also compose with the order properties of sequences. In the simplest case, cycling (see above) establishes a rhythmic subset and a definitive order relation among the cyclic grouping.

Beyond this simple procedure we move into the whole area of the manipulation of finite sets of entities which has been very extensively researched and used in late Twentieth Century serial composition. Starting originally as a method for organising Hpitches which have specific relational properties due to the cyclical nature of the pitch domain (see Chapter 13) it was extended first to the time-domain, which has some similar properties, and then to all aspects of notated composition.

Permutation of a finite set of elements in fact provides a way of establishing relationships among element-groupings which, if the groupings are small, are clearly audible. If the sets are large, such permutations preserve the field (reference-set) so that complex musical materials are at least field-related (claims that they are perceptually order-related are often open to dispute). All the insights of serial or permutational thinking about materials may be applied to any type of sequence, provided that we ask the vital question, "is this design audible?". And in what sense is it audible? – as an explicit re-ordering relation – or as an underlying field constraint?

In traditional scored music, the manipulation of set order-relations is usually confined to a subset of properties of the sounds e.g. Hpitch, fixed duration values, a set of loudness values. But we can work with alternative properties, or groups of properties, such as total spectral type. For example, permutations of instrument type in a monodic klangfarbenmelodie – Webern's Opus 21 Sinfonie provides us with such an instrument-to-instrument line and uses it canonically. We could, however, permute the instrumental sequence and not the Hpitches. Formant sequence e.g.

bo–ba–be = do–da–de = lo–la–le

and onset characteristic sequence e.g.

bo–ka–pe = bu–ki–po = ba–ko–pu etc.

can be most easily observed as patternable properties in language or language-derived syllables, but formant and onset characteristics can be extracted and altered in all sounds.

We might imagine an unordered set of sounds and which define no small reference set, on which we impose a sequence of "bright", "hard", "soft", "subdued" etc. attacks as a definable sequence. The continuations of such sounds might be recognisable (e.g. water, speech, a bell, a piano concerto excerpt etc.) but ordering principles are applied just to the onset types. This is an extreme example because I wish to stress the point that order can be applied to many different properties and can be used to focus the listener's attention on the ordered property, or to contrast the area of order manipulation with that of lesser order. In the setting of prose this contrast is clear. In our extreme example we have suggested a kind of ordering of onset characteristics laid against a "cinematographic" montage which might have an alternative narrative logic – the two might offset and comment upon one another in the same way that motivic (and rhythmic) ordering and the narrative content of prose do so in traditional musical settings of prose.

Ordering principles can, of course, also be applied to sounds as wholes. So any fixed set of sounds can be rearranged as wholes and these order relations explored and elaborated. A traditional example is English change-ringing where the sequence in which a fixed set of bells is rung goes through an ordered sequence of permutations. (See **Diagram 1**). (**Sound example 7.18**).

STRESS PATTERNS AND RHYTHM

We can apply different order-groupings to mutually exclusive sets of properties (e.g. onset, continuation, loudness, as against Hpitch) within the same reference-set of sounds. Onset characteristics and overall loudness are often organised cooperatively to create stress patterns in sequences. Combined with the organisation of durations, we have the group of properties used to define rhythm. Rhythmic structure can be established independently of Hpitch structure (as in a drum-kit solo) or parallel to and independently of Hpitch order relations. Most pre twentieth century European art music and e.g. Indian classical music, separates out these two groupings of sound properties and orders them in a semi-independent but mutually interactive way (HArmonic rhythm in Western Music, melodic/rhythmic cadencing in Indian Classical Music).

Again, whole volumes might be devoted to rhythmic organisation but as there is already an extensive literature on this, I will confine myself to just a few observations.

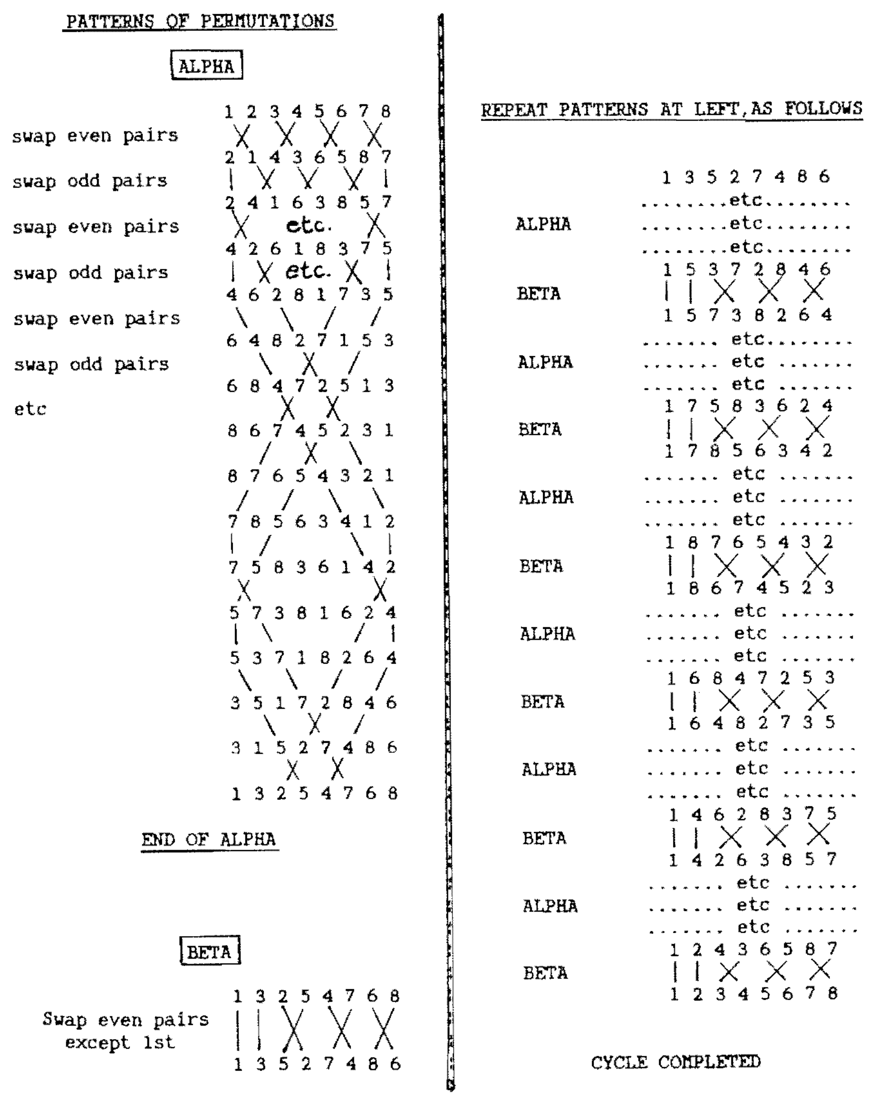
Stress patterns may be established, even cyclical stress patterns, independent of duration organisation, or on a regular duration pattern that is subsequently permuted or time-warped in a continuously varying manner, a feature over which we can have complete control in sound composition. Hence, stress duration patterns at different tempi, or at different varying tempi, can be organised in such a way that they synchronise at exactly defined times, these times perhaps themselves governed by larger time-frame stress patterns. (See Chapter 9). (**Sound example 7.19**).

By subtle continuous control of relative loudness or onset characteristics, we can change the perception of the stress pattern. It might, for example, be possible to create superimposed grouping perceptions on the same stream of semiquaver units, e.g. grouped in 5 and in 7 and in 4 and to change their relative perceptual prominence by subtly altering loudness balance or onset characteristics or in fact to destroy grouping perception by a gradual randomisation of stress information. Here field variation is altering order perception. (**Sound example 7.20**).

DIAGRAM 1

ENGLISH BELL-RINGING PATTERNS

PLAIN BOB



Alternatively, Clarence Barlow has defined an "indispensability factor" for each event in a rhythmic grouping, e.g. a 6/8 bar. This defines how important the presence or absence of an event is to our perception of that grouping. In 6/8 for example, the first note of the 6 is most important, the 4th note (which begins the 2nd set of 3 notes) the next, and so on. In 3/4 over the same 6 quavers, the indispensability factors will be arranged in a different order, stressing the division of the 6 into 3 groups of 2. These factors are then used to define the probability that a note will occur in any rhythmic sequence and by varying these probabilities, we can e.g. vary our perception of 6/8–ness versus 3/4–ness. Also, by making all notes equally probable, all stress regularity is lost and the sequence becomes arhythmic above the quaver pulse level. **(Diagram 2)**. **(Sound example 7.21)**.

ORDERING ORDER

We can also work with order–groupings of order–groupings e.g. with Hpitches, altering the time–sequencing of motivic–groups–taken–as–wholes. There is an extensive musical literature discussing order–regrouping, especially of Hpitches, and all such considerations can be applied to any subset of the properties of sounds, or to sounds as wholes.

Here we are beginning to stray beyond the frame of reference of this chapter, because reordering principles can be applied on larger and larger time–frames and are the substance of traditional musical "form". From a perceptual point of view, I would argue that the larger the time–frame (phrase–frame and beyond), the more perception is dependent on long term memory, and the less easy pattern–retention and recall becomes. Hence larger scale order relations tend to become simpler, in traditional practice, as we deal with larger and larger time–frames. This is not to say that, in traditional practice, the "repeated" objects are not *internally* varied in more complicated ways at smaller time–frames, but as large time–block qua large time–blocks, the permutation structure tends to be simpler. (Such matters are discussed in greater detail in Chapter 9).

The computer, having no ear or human musical judgement, can manipulate order sequences of any length and of any unit–size in an entirely equivalent manner. As composers, however, we must be clear about the relationships of such operations to our time experience. Equivalence in the numerical domain of the computer (or for that matter on the spatial surface of a musical score) is not the same as experiential equivalence. Audible design requires musical judgement and cannot be "justified" on the basis of the computer–logic of a generating process.

In fact, as the elements of our sequence become larger, we pass over from one time–frame to another. Thus the temporal extension of a sequence (by unit respacing, *time–stretching* etc.) is a way of passing from one perceptual time–frame to another, just as time–shrinking will ultimately contract the sequence into an essentially continuous perceptual event. With time–expansion, the perceptual boundaries are less clear, but no less important.

Permutational logic is heady stuff and relationships of relationships can breed in an explosive fashion, like Fibonacci's rabbits. Permuting and otherwise reordering sequences in ever more complicated ways is something that computers do with consummate ease. A powerful order–manipulation instrument can be written in a few lines of computer code, whereas what might appear to be a simple spectral–processing procedure might run to several pages. The questions must always be, does the listener hear those reorderings and if so, are they heard directly, or as field constraints on the total

DIAGRAM 2

experience? If the latter, need they be quite so involved as I suppose, i.e. is there a simpler, and hence more elegant, way to achieve an equivalent aural experience? A more difficult question, as with any musical process is, does it matter to the listener that they hear this reordering process? Or, what does this reordering "mean" in the context of an entire composition? Is it another way of looking down a kaleidoscope of possibilities, or a special moment in a time-sequence of events taking on a particular significance in a musical-unfolding because of its particular nature and placement with respect to related events? These are, or course, aesthetic questions going beyond the mere use, or otherwise, of a compositional procedure.

CHAPTER 8

TEXTURE

WHAT IS TEXTURE ?

So far in this book we have looked at the intrinsic or internal properties of sounds. However in this chapter we wish to consider properties of dense agglomerations of the same or similar (e.g. *tape-speed variation* transposed) versions of the same sound, or set of sounds. This is what I will call texture. Initially we will assume that what I mean by texture is obvious. In the next chapter we will analyse in more detail the boundary between textural and measured perception, particularly in relation to the distribution of events in time.

In the first two sound examples, we give what we hope will be (at this stage) indisputable examples of measuredly perceived sound and texturally perceived sound. In the first we are aware of specific order relations amongst Hpitches and the relative onset time of events and can, so to speak, measure them, or assign them specific values (this will be explained more fully later). In the latter we hear a welter of events in which we are unaware of any temporal order relations. However, in the latter case, we are aware that the sound experience has some definable persisting properties – the Hpitches define a persisting reference set (a HArmonic field). **(Sound example 8.1)**.

We can lose the sense of sequential ordering of a succession of sounds in two ways. Firstly, the elements may be relatively disordered (random) in some property (e.g. HPitch). Secondly the elements may succeed each other so quickly that we can no longer grasp their order.

There are immediate perceptual problems with the notion of disorder. We can generate an order-free sequence in many ways, but it is possible to pick up on a local ordering in a totally disordered sequence. Thus a random sequence of zeros and ones may contain the sequence 11001100 which we may momentarily perceive as ordered, even if the pattern fails to persist. Such focusing on transient orderliness may be a feature of human perception as we tend to be inveterate pattern-searchers. So disorderly sequence, of itself, need not lead to textural perception. **(Sound example 8.2)**.

By the same token, if a sequence, no matter how rapid, is repeated (over and over), the sequence shape will be imprinted in our memory. This 'looping effect' can thus contradict the effect of event-rate on our perception. **(Sound example 8.3)**.

Textural perception therefore only takes over unequivocally when the succession of events is both random and dense, so we no longer have any perceptual bearings for assigning sequential properties to the sound stream. **(Sound example 8.4)**.

A disordered sequence of Hpitches in a temporally dense succession is a fairly straightforward conception. However, we can also apply the notion of texture to temporal perception itself. This involves more detailed arguments about the nature of perception and we will leave this to the next chapter.

So, broadly speaking, texture is sequence in which no order is perceived, whether or not order is intended or definable in any mathematical or notatable way. Texture differs from Continuum in that we retain a sense that the sound event is composed of many discrete events. Pure textural perception takes over from measured perception when we are aware only of persisting field properties of a musical stream and completely unaware of any ordering properties for the parameters in question.

In some sense, texture is an equivalent of noise in the spectral domain, where the spectrum is changing so rapidly and undirectedly that we do not latch onto a particular spectral reference frame and we hear an average concept, "noise". But, like noise, texture comes in many forms, has rich properties and also vague boundaries where it touches on more stable domains.

GENERATING TEXTURE STREAMS

The most direct way to make a texture-stream is through a process that *mixes* the constituents in a way given by higher order (density, random scatter, element and property choice) variables. There are many ways to do this and we will describe just three.

We may use untransposed sound-sources, specifying their timing and loudness through a *mixing score* (or a graphic mixing environment, though in high density cases this can be less convenient) and use various *mix shuffling* meta-instruments to control timing, loudness and sound-source order.

Alternatively textural elements may be submitted as 'samples' (on a 'sampler') or, equivalently, as sampled sound in a look-up table for submission to a table-reading instrument like *CSound*. Textures can then be generated from a MIDI keyboard (or other MIDI interface), using MIDI data for note-onset, note-off, key velocity, and key-choice to control timing, transposition and loudness (or, in fact, any texture parameters we wish to assign to the MIDI output) – or from a CSound (textfile) score.

The keyboard approach is intuitively direct but difficult to control with subtlety when high densities are involved. The CSound score approaches requires typing impossible amounts of data to a textfile, but this can be overcome by using a meta-instrument which generates the CSound score (and 'orchestra') from higher level data.

Via such *texture control* or *texture generation* procedures (see **Appendix pp68–69**), we can generate a texture for a specified duration from any number of sound-sources, giving information on temporal-density of events, degree of randomisation of onset-time, type (or absence) of event-onset quantisation (the smallest unit of the time-grid on which events must be located), pitch-range (defined over the continuum or a prespecified, possibly time-varying, HArmonic field), range of loudness from which each event gets a specific loudness, individual event duration, and the spatial location and spatial spread of the resulting texture-stream. In addition all of these parameters may vary (independently) through time.(see **Appendix pp68–69**). **(Sound example 8.5)**.

Thirdly, any shortish sound may be used as the basis for a texture-stream, if used as an element in *granular synthesis* where the onset time distribution is randomised and the density is high, but not extremely high. The properties of the sound (loudness trajectory, spectral brightness etc) may be varied from unit to unit to give a diversity of texture-stream elements. **(Sound example 8.6)**.

Alternatively we may begin with individually longer sound sources. Grain streams and sequences may begin to take on texture-stream characteristics by becoming arhythmic (through the scattering of onset-time away from any time reference-frame: see later in this Chapter) and by becoming sequentially disordered. If we now begin to superimpose variants, even with a few superimpositions we will certainly generate a texture stream. **(Sound example 8.7)**.

Continuous sounds which are evolving through time (pitch-glide, spectral change etc) may also change into texture-streams. Applying *granular reconstruction* (see **Appendix p73**) to such a continuous sound with a grain-size that is large enough to retain perceptible internal structure, and provided both that the density is not extremely high (when the sound becomes continuous once again) and the grain time distribution is randomised, can produce a texture stream. **(Sound example 8.8)**.

Spatialisation can be a factor in the emergence of a texture-stream. If a rapid sequence has its elements distributed alternately between extreme left and right positions, the percept will be most probably split into two separate sequences, spatially distinct and with lower event rates. However, if the individual events are scattered randomly over the stereo space, we are more likely to create a stereo texture-stream concept as the sense of sequential continuity is destroyed, particularly if the onset-time distribution is randomised. A superimposition of two such scattered sequences will almost certainly merge into a texture-stream. **(Sound example 8.9)**.

And in fact any dense and complex sequence of, possibly layered, musical events, originally having sequential, contrapuntal and other logics, can become a texture-stream if the complications of these procedures and temporal density are pushed beyond certain perceptual limits.

FIELD

The two fundamental properties of texture are Field and Density.

Field refers to a grouping of different values which persists through time. Thus a texture may be perceived to be taking place over a whole-tone scale, even though we do not retain the exact sequence of Hpitches. In this case, we retain a HArmonic percept. Similarly, we may be able to distinguish French from Portuguese, or Chinese from Japanese, even if we do not speak these languages and hence do not latch onto significant sequences in the speech stream. This is possible because the vowel-formants, consonant-types, syllabic-combinations or even the pitch-articulation types for one language form a field of properties which are different from those of another language.

Even aspects of the time organisation may create a field percept. Thus, if the time-placement of events is disordered so that we perceive only an indivisible agitation over a rhythmic stasis, but this placement is quantised over a very rapid pulse (e.g. events only occur on time-divisions at multiples of 1/30th of a second), we may still be aware of this regular time-grain to the texture. This is a field percept. A more complete discussion of temporal perception can be found in the next chapter. **(Sound example 8.10)**.

Field properties themselves may also be evolving through time, in a direct or oscillating manner. One method of creating sound *interpolation* (see Chapter 12) is through the use of a texture whose elements are gradually changing in type. In **Sound Example 8.11** we hear a texture of vocal utterances which becomes very dense and rises in pitch. As it does so, its elements are *spectrally-traced* (see **Appendix p25**) so that the texture-stream becomes more pitched in spectral quality.

There are four fundamental ways to change a field property of a texture-stream. (see **Diagram 1**).

(a) We may gradually move its value.
(b) We may spread its value to create a range of values.
(c) We may gradually change the range of values.
(d) We may shrink the range to a single value.

For example, we may reduce the event duration so that the texture might become grittier (this also depends on the onset characteristics of the grains). We might begin with a fixed spectral type defined mainly by *enveloping* (see Chapter 9) or by source class (e.g. twig snaps, pebbles on tiles, bricks into mud, water drips etc.) then *spectrally interpolate* (see Chapter 12) over time from one to another, or gradually expand the spectral range from just one of these to include all the others. We may begin with pitched spectra and change these gradually to inharmonic spectra of a particular type, or to inharmonic spectra of a range of types, or noisy spectra etc. We may begin with a fixed-pitch and gradually spread events over a slowly widening pitch-band (*wedgeing* : **Appendix p69**).

If our property has a reference-frame (see Chapter 1), e.g. an Hpitch set for pitches, these field properties of the texture may gradually change in a number of distinct ways.

In terms of an Hpitch field.....

(a) We may change the Hpitch field by *deletion*, *addition* or *substitution* of field components. Addition and substitution imply a larger inclusive Hpitch field and the process is equivalent to a HArmonic change over a scale system (e.g. the tempered scale) in homophonic tonal music. (see **Diagram 2a**).

(b) We may change the Hpitch field by *gradually retuning* the individual Hpitch values towards new values. This kind of gradual tuning shift can destroy the initial Hpitch concept and reveals the existence of the pitch continuum. (see **Diagram 2b**).

(c) We may change the Hpitch field by *spreading* the choice of possible values of the original Hpitches. Each Hpitch can now be chosen from an increasing small range about a median value. Initially we will retain the sense of the original Hpitch field, but eventually, as the range broadens, this will dissolve into the pitch continuum. (see **Diagram 2c**).

(d) We may gradually destroy the Hpitch field by *gliding* the pitches. Once the gliding is large, and not start- or end- focused (see Chapter 2), we lose the original Hpitch percept. (See **Diagram 2d**).

Such reference-frame variations can be applied to other properties which are perceived against a reference frame. (**Sound example 8.12**).

DIAGRAM 1

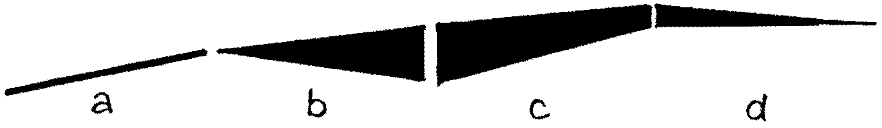
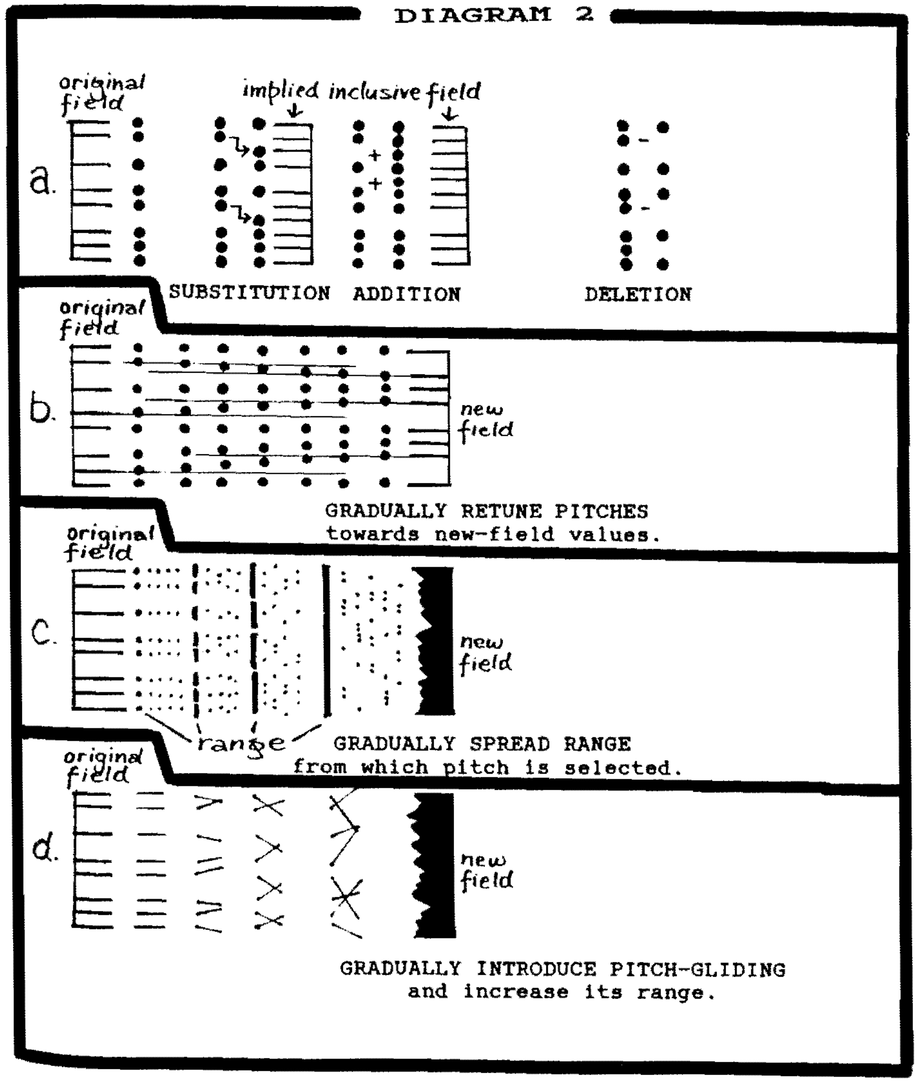


DIAGRAM 2



We cannot hope to describe all possible controllable field parameters of all texture-streams because, as the elements of the stream are perceptible, we would need to describe all possible variations of each constituent, all combinations of these properties, all changes of the individual properties and all changes in combinations of these properties. We will therefore offer a few other suggestions.......

(1) Variation of sound-type of the constituents (harmonicity-inharmonicity; type of inharmonicity; formant-type; noisiness-clarity; stability or motion of any of these); the range of sound-types; the temporal variation of these.

(2) Variation of the individual spatial motion of elements; the temporal variation of this.

And, if the texture elements are groups of smaller events...

(3) The group size and its variation; the range of group size and its variation.

(4) The internal pitch-range, spectral-range (various) of the groups.

(5) The internal group-speed, group-speed range and their slow or undulating variation.

(6) The internal spatialisation of groups (moving left, spatial oscillation, spatial randomness), range of spatialisation types, and the time-variation of these.

(7) The variation of order-sequence or time-sequence of the groups.

All such features may be compositionally controlled independently of the stream density and the onset-time randomness of the texture-stream.

DENSITY

The events in a texture-stream will also have a certain density of event-onset-separation which we cannot measure within perception but which we can compare with alternative densities. Thus we will be able to perceive increases and decreases in density, oscillations in density and abrupt changes in density. We have this comparative perception of density changes so long as these changes are in a time-frame sufficiently longer than that of density perception itself. Otherwise there is no way to distinguish density from density fluctuation. (**Sound example 8.13**).

In fact event-onset-separation-density perception is like temperature measurement in a material medium. Temperature is a function of the speed and randomness of motion of molecules. Hence the temperature at a point, or of a single molecule has no meaning. In the same way, density at the time-scale of the event repetitions has no meaning. Temperature can only be measured over a certain large collection of molecules - and density can only be measured over a group of event entries. Spatial variation in temperature similarly can only be measured over an even larger mass - and temporal variation in density similarly requires more events, more time, to be perceived than does density itself.

Event-onset-separation-density fluctuations themselves may be random in value but regular, or semi-regular in time, providing a larger time-frame reference-frame. Furthermore, slow density fluctuations will tend to be perceived as directed flows - as continuation properties - as compared with

rapid density fluctuations. In extremely dense textures, the density fluctuations themselves may approach the size of large grains - we create granular or "crackly" texture. (**Sound example 8.14**).

Increasing the density of a complex set of events can be used to average out its spectral properties and make it more amenable to interpolation with other sounds. In fact, before the advent of computer technology this was one of the few techniques available to achieve sound *interpolation* (see Chapter 12).

In **Sound example 8.15** (from *Red Bird*: 1977, a pre-digital work), a noisy-whistled voiced is interpolated with a (modified) skylark song via this process of textural thickening. In **Sound example 8.16** (from *Vox 5* : 1986), behind the stream of complex vocal multiphonics, we hear the sound of a crowd emerging from a vocal-like noise band which is itself a very dense texture made from that crowd sound.

When this process is taken to its density extreme we can produce *white-out*. When walking on snow in blizzard conditions it is sometimes possible to become completely disoriented. If the snow fall is sufficiently heavy, all sense of larger outlines in the landscape are lost in the welter of white snow particles. Similarly, when a sound texture is made extremely dense, we loose perceptual track of the individual elements and the sound becomes more and more like a continuum. In particular, where the sound elements are of many spectral types including noise elements, the texture goes over into noise. This is what we describe as white-out.

In **Sound example 8.17**, a dense texture of vocal sounds whites-out and the resulting noise-band is then *filtered* to give it spectral pitches while it slides up and down. The pitchness is then gradually removed from all bands except one, and the density decreases once more to reveal the original human voice elements.

Finally we should note that changes in field and density properties might be coordinated between parameters, or all varied independently, or even have nested 'contradictory' properties. Thus a texture may consist of events whose onsets are entirely randomly distributed in time but whose event-elements are clearly rhythmically ordered *within* themselves, or, conversely, events may begin on the Hpitches of a clearly defined HArmonic field while the event-elements have Hpitches distributed randomly over the continuum. (**Sound example 8.18**).

CHAPTER 9

TIME

ABOUT TIME

In this chapter we will discuss various aspects of the organisation of time in sound composition. Clearly, rhythm is a major aspect of such a discussion but we will not dwell on rhythmic organisation at length because this subject is already dealt with in great detail in the existing musical literature. We will be more concerned with the nature of rhythmic perception and its boundaries.

Our discussion will enable us to extend the notion of perceptual time-frames to durations beyond that of the grain and upwards towards the time-scale of a whole piece.

WHERE DOES DENSITY PERCEPTION BEGIN ?

In the previous chapter we discussed texture-streams which were *temporally* dense but which might retain field properties in other dimensions (like Hpitch or formant-type). We must now admit that the concept of density and density variation can be applied to any sound parameter. For example, if we have pitch confined over a given range, a pitch density value would tell us how densely the pitch-events covered the continuum of pitch values between the upper and lower limits of the range (not time-wise but pitch-wise).

In this case we can begin to see that the concepts of Density and Field applied over the *same* parameter come into conflict. Once the pitch-density (in this new sense) becomes very high, we loose any sense of a specific Hpitch field or HArmonic reference frame, though we may continue to be aware of the range limits of the field. **(Sound example 9.1)**.

We must therefore ask, what is the dividing line between field, or reference-frame, perception and density perception *in any one dimension*. In this chapter we will confine ourselves to the dimension of temporal organisation. Our conclusions may however be generalised to the field/density perceptual break in any other dimension (e.g. Hpitch organisation).

Compositionally, we can create sequences of events that gradually lose a (measured) sense of rhythm. Thus we may begin with a computer-quantised rhythmic sequence which has an "unnatural" or "mechanical" precision. Adding a very small amount of random scatter to the time-position of the events, gives the rhythm a more "natural" or "human performed" feel. This is because very accurately performed rhythmic music is not "accurate" in a precisely measured sense but contains subtle fluctuations from "exactness" which we regard as important and often essential to a proper rhythmic "feel".

Increasing the random scatter a little further we move into the area of loosely performed rhythm, or even badly performed rhythm and eventually the rhythm percept is lost. The time-sequence is arhythmic. Once this point is reached we perceive the event succession as having a certain density of event onsets – our perception has changed from grasping rhythmic ordering as such to grasping only density and density fluctuations. **(Sound example 9.2)**.

In this sequence we probably retain the sense of an underlying measured percept (which is being articulated by random scattering) a long way into the sequence. We are given a reference frame by the initial strict presentation, which we carry with us into the later examples. If we were presented with some of the later examples *without* hearing the reference frame, we would perhaps be more willing to declare them arhythmic.

Taking the sequence in the opposite order, there may be a perceptual switching point at which we suddenly become aware of rhythmic order in the sequence. **(Sound example 9.3)**.

Let us now look at this situation from another viewpoint. Beginning again with our strictly rhythmic set of events, we note that the event-onsets lie on (or very close to) a perceivable time grid or reference-frame (the smallest common beat subdivision, which may also be thought of as a time-quantisation grid). Allowing event-onsets to be displaced randomly by very small amounts from this time reference-frame, we initially retain the percept of this reference frame and of a rhythm in the event stream. Once these excursions are large, however, our perception of the frame, and in consequence rhythmicity, breaks down.

It is informative to compare this with the analogous situation pertaining to an Hpitch reference frame. Here we would begin with events confined to an Hpitch set (a HArmonic field), then gradually randomise the tuning of notes from the Hpitch set, slowly destroying the Hpitch field characteristics, even though we might retain the relative up-downness in the pitch sequencing.

From this comparison we can see that a durational reference-frame underlying rhythmic perception is similar to a field, and rhythm is an ordering relation over such a reference frame. Dissolving rhythmicity is hence analogous to dissolving the percept of Hpitch which also relates intrinsically to a reference set.

Strictly speaking, to provide a precise analogy with our use of HArmonic field, a duration field would be the set of all event-onset-separation durations used in a rhythmic sequence. However, just as underlying any HArmonic field we may be able to define a frame made up of the smallest common intervallic unit (e.g. the semitone for scales played in the Western tempered scale, the srutis of the Indian rag system), it is more useful to think of the smallest subdivision of all the duration values in our rhythmic sequence, which we will refer to as the time-frame of the event. In an idealised form, this may also be thought of as the time quantisation grid.

Such a time-frame, constructed from our perception of event-onset-separation duration, provides a perceptual reference at a particular scale of temporal activity. As such it provides us with a way to extend the notion of perceptual time-frames used previously to define sample-level, grain-level and continuation-level perception (or lack of it) into longer swathes of time. Moreover, because such time-frames may be nested (see below) we can in fact define a hierarchy of time-frames up to and including the duration of an entire work.

Just as with the dissolution of Hpitch perception, it is the dissolution of the time-frame which leads us from field-ordered (rhythmic) perception of temporal organisation to density perception. And just as dissolving the Hpitch percept by the randomisation of tuning leaves us with many comparatively perceived pitch properties to compose, dissolving the time-reference-frame leads us into the complex domain of event-onset-separation-density-articulation discussed in the previous chapter.

MEASURED, COMPARATIVE & TEXTURAL PERCEPTION

To make these distinctions completely clear we must examine the nature of time-order perception and define our terms more precisely.

In the ensuing discussion we will use the term 'duration' to mean event-onset-separation-duration, to make things easier to read. Bear in mind however that duration, here, means only this, and not the time-length of sounds themselves.

There are two dimensions to our task. Firstly, we must divide temporal perception into three types – the measured, the comparative and the textural (the perception of temporal density and density fluctuation).

Secondly, there is the question of time-frames. As will be discussed below, time-frames may be nested within one another in the rhythmic or temporal organisation of a piece so that at one level (e.g. a semiquaver time-frame) order-sequences are put in motion (rhythm) while simultaneously establishing a time-frame at the larger metrical unit (e.g. the bar) in which longer-duration order-sequences can be set up.

So we can ask, in what time-frames is our perception of a particular event measurable and in what time-frames is it comparative or textural ?

The first assertion we will make is that measured perception requires an established or only slowly changing reference-frame against which we can "measure" where we are. In the pitch domain this might be a tuning system, or a mode or scale, or, in Western tonal music, a subset of the scale defining a particular key. This permits structural judgements like "this is an E-flat" or "we are in the key of C# minor". Such a reference-frame might be established by cultural norms (e.g. tempered tuning) or established within the context of a piece (e.g. the initial statement of a raga, the establishment of tempo and metrical grouping in a particular piece). If we do not have a reference set, we are still able to make comparative judgements. In the pitch domain for example, "we are now higher than before", or, "we are moving downward", or, "we are hovering around the central value".

In the sphere of durations, if we establish a time reference-frame of, say, repeated quavers at [crotchet = 120] we can recognise with a fair degree of precision the various multiples (dotted crotchet, minim, etc) and integral divisions (semiquaver, demisemiquaver, triplet semiquaver) of this unit. (See **Diagram 1**). Hence we can, in many cases, measure our perception against the reference frame and recognise specific duration sequences. Our perception of duration is measured, at the time-frame of the quaver.

Similarly in the sequence in **Diagram 2** we will probably hear a clearcut sequence of duration values in the perceptibly measured ratios 1:2:3:4:5. Our perception is again measured.

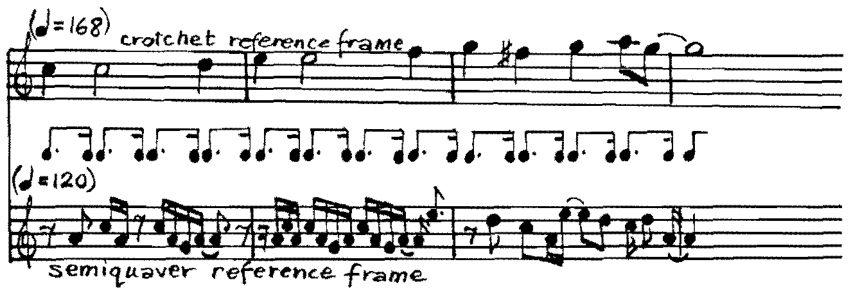
However, in the situation in **Diagram 3** we are aware of the comparative quality of the successive durations. We perceive them as getting relatively longer and we also have an overall percept of "slowing down" but we do not (normally) have a clear percept of the exact proportions appertaining between the successive events. Furthermore no two live performances of this event will preserve the same set of measured proportions between the constituents. This is, then, an example of comparative perception.

DIAGRAM 1

DIAGRAM 2

DIAGRAM 3

DIAGRAM 4



A more interesting example is presented by the sequence in **Diagram 4**. If this rhythm occurs in a context where there is a clear underlying semiquaver reference-frame, and the sequence is played "precisely" as written, we will perceive the 3:1:3:1:3:1 etc. sequence of duration proportions clearly – our perception will be measured. However, in many cases, this time pattern is encountered where the main reference frame is the crotchet, and the pattern may be more loosely interpreted by the performer, veering towards 2:1:2:1: etc. at the extreme. Here we are perceiving a regular alternation of short and long durations which, however, are not necessarily perceived in some measurable proportions. The score may given an illusory rigour to a 3:1 definition, but we are concerned here with the percept. **(Sound example 9.4)**.

The way in which such crotchet beats are divided is one aspect of a sense of "swing" in certain styles of music. A particular drummer, for example, may have an almost completely regular long–short division of the crotchet in the proportion 37:26. We may be aware of the regularity of his/her beat and appreciative of the particular quality of this division in the sense of the particular sense of swing it imparts to his/her playing while remaining completely unaware of the exact numerical proportions involved. Here then we have a comparative perception with fundamental qualitative consequences. **(Sound example 9.5)**.

It is important to note at this point that our perception of traditional fully-scored music is comparative in many respects and in some respects textural to the extent that e.g. the precise morphology of each violin note cannot be specified in the notation and varies arbitrarily over a small range of possibilities, as does the vibrato of opera singers. In sound composition, these factors can be precisely specified and, if desired, raised to the level of comparative, or measured, perception.

More importantly, our rhythmic example using dotted rhythms illustrates the second aspect of our discussion. For in this example, perception at the level of the crotchet remains measured – the music is "in time". However, simultaneously, in a smaller time-frame, our perception has become comparative.

We can see the same division into time-frames if we look again at the idea of "indispensability factor" proposed by Clarence Barlow (see Chapter 7). As discussed previously, we can define, over a reference frame of quavers, the relative indispensability of each note in a 6/8 or a 3/4 (or a 7/8) pattern. Linking the probability of occurrence of a note to its indispensability factor allows us to generate a strong 6/8 grouping feel, or a strong 3/4 feel, or an ambiguous percept between the two. Once every quaver becomes equally probable, however, the sense of grouping breaks down altogether and at the level of 6-groupings (7-groupings, or any groupings) measured perception is lost. Our perception reverts to the field characteristics. In this case, however, the field is defined by a smaller set of durations, the quavers themselves. So at the smaller time-frame, we retain a sense of measured regularity and hence our perception *there* is measured perception.

SHORT DURATION REFERENCE-FRAMES : RHYTHMIC DISSOLUTION

There are, however, limits to this time-frame switch phenomena. On the large scale, if the pulse cycle becomes too long (e.g. 30 minutes), we will no longer perceive it as a time reference-frame (some musicians will dispute this, see below). More significantly, if the pulse-frame becomes too small we also lose a sense of measurability and hence of measured perception. We cannot give a precise figure for this limit. We can perceive the regularity of grain down to the lower limit of grain perception but comparative judgements of grain-durations, especially in more demanding proportions (5:7 as opposed to 1:2), seems to break down well above this limit.

DIAGRAM 5

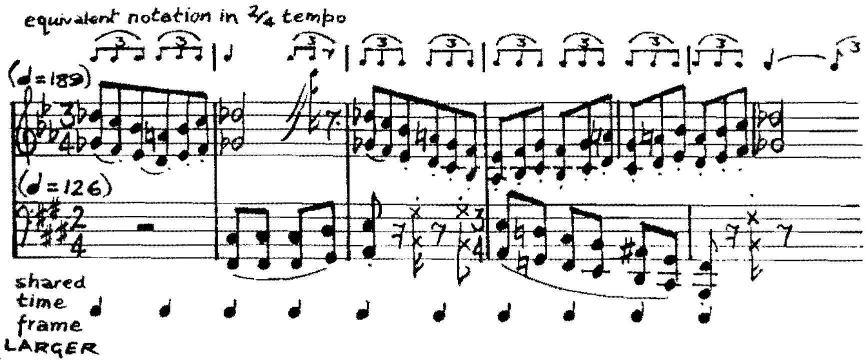
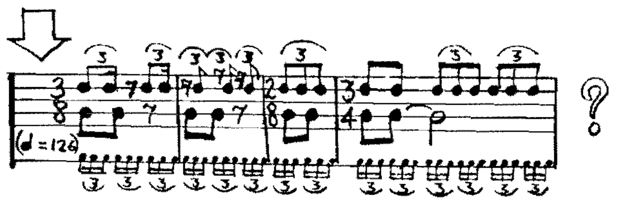


DIAGRAM 6

DIAGRAM 7



This problem becomes particularly important where different divisions of a pulse are superimposed. To given an example, if we compose two duration streams in the proportion 2:3 (see **Diagram 5**) we have a clear mutual time-frame pulse at the larger time-frame of the crotchet. (**Sound example 9.6**).

If we now regroup the elements in each stream in a way which contradicts this mutual pulse (see **Diagram 6**) we may still be able to perceptually integrate the streams (perceive their exact relationship) in a measured way over a smaller time-frame. By dividing the quavers of the slower stream into 3 and the faster stream into 2, we may discover (i.e. perceive) a common pulse. (see **Diagram 7**). (**Sound example 9.7**).

Even here this perception does not necessarily happen (certainly not for all listeners), particularly where we are relying on the accuracy of performers.

The more irrational (in a mathematical sense) the tempo relationship between the two streams, the shorter this smaller mutual time-frame pulse becomes. And the smaller this unit, the more demanding we must be on performance accuracy for us to hear this smaller frame. With computer precision, however, this underlying mutual pulse may continue to be apparent in situations where it would be lost in live performance, as in **Diagram 8**. (**Sound example 9.8**).

Perceiving a *specific* 9 to 11 proportion as in the last example, where the smaller common time-frame unit has a duration of c. 100th of a second, is simply impossible, even given computer precision in performance. Measured perception on the smaller mutual time-frame has broken down, though we may be comparatively aware (if the density of events-relative-to-the-stream-tempo can be compared in the two streams) that we have 2 streams of similar but different event density.

With 3 superimposed tempi the problem is, of course, compounded, the common pulse unit becomes even shorter. We may, for example, when composing on paper, set up sequences of events in which simultaneous divisions of the crotchet beat into (say) 7,11, & 13 at [crotchet = 120], are used. (See **Diagram 9**).

And we may always *claim* that we are setting up an exact percept created by the exact notational device used. However, in this particular case we should ask, with what accuracy can this concept be realised in practice?

No human performance is strictly rhythmically regular (see discussion of quantised rhythm above). When three streams are laid together the deviations from regularity will not flow in parallel (they will not synchronise like the parallel micro-articulations of the partials in a single sound-source: see Chapter 2). We may describe the fluctuations of the performed lines from exact correlation with the common underlying pulse by some scattering factor, a measure of how much an event is displaced from its "true" position. A factor 1 means it is displaced by a whole unit. In larger time-frame terms, a quaver in a sequence of quavers would be inaccurately placed by up to a whole quaver's length. By anyone's judgement this would have to be described as an inaccurate placement! (See **Diagram 10**).

In fact I would declare the situation in which events are randomly scattered within a range reaching to half of the duration of the time-frame units as definitively destroying the percept of that time-frame. In practice the time-frame percept probably breaks down with even more closely confined random displacements. (**Diagram 11**).

DIAGRAM 8

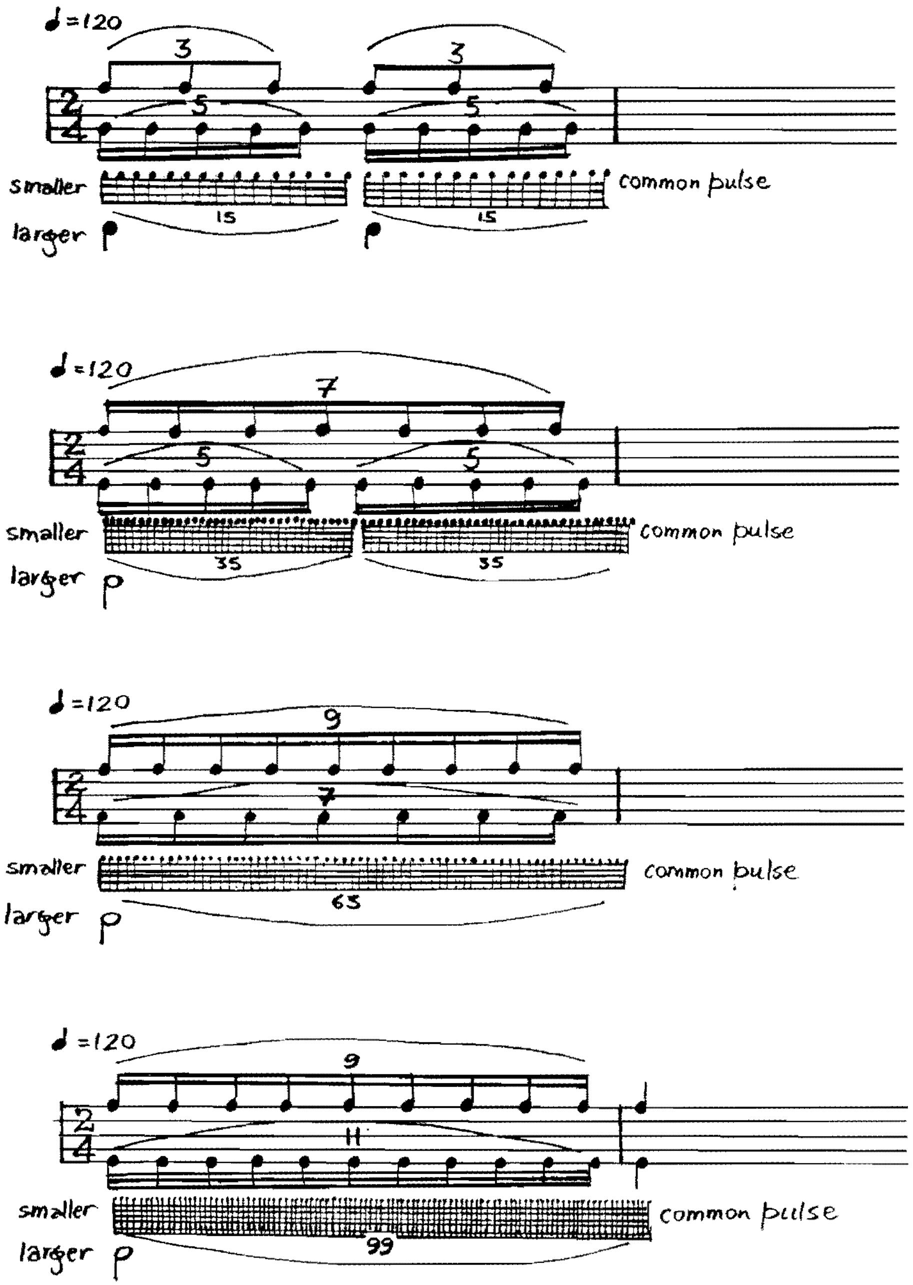


DIAGRAM 9

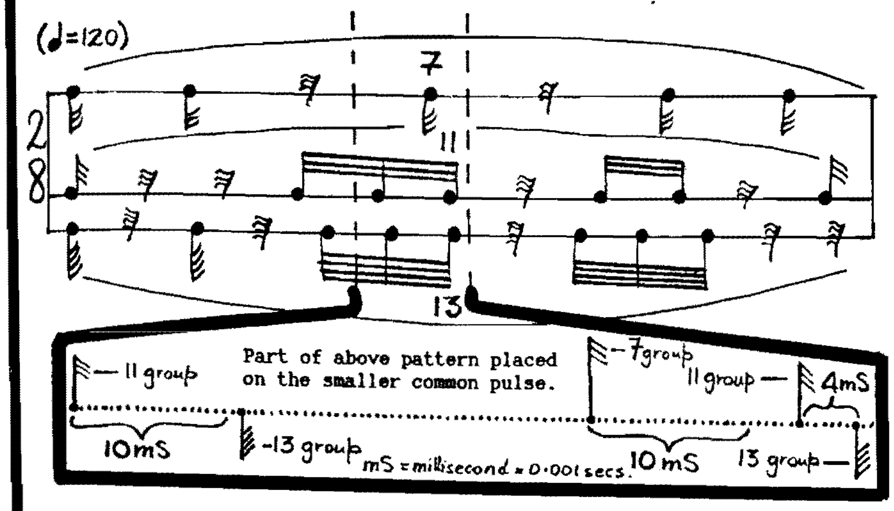


DIAGRAM 10

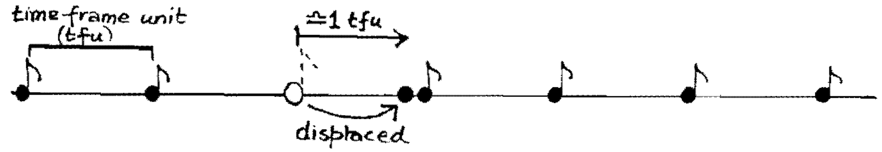


DIAGRAM 11

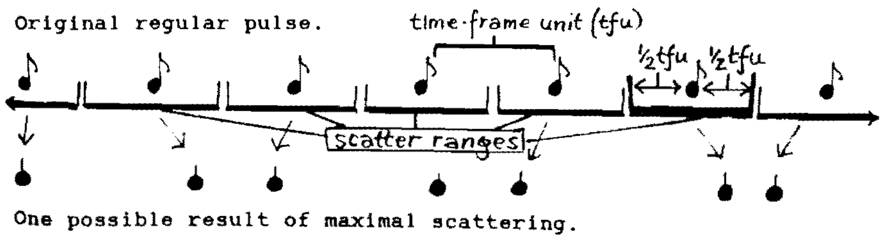
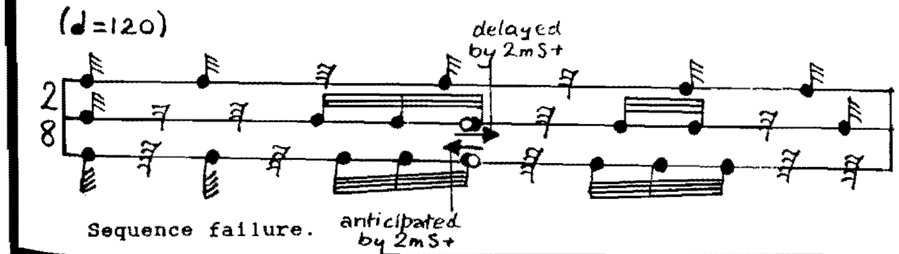


DIAGRAM 12



In our notated example, the mutual time-frame duration lasts approximately 0.0005 seconds, or half a millisecond, and hence there is no doubt that the live-performed events will be displaced by at least half the time-frame unit (half of half a millisecond). In fact we can confidently declare that events will be displaced by multiples of the time-frame unit, in all performances. The precision of the result is simply impossible and an appeal to future idealised performances mere sophistry.

In fact we can be fairly certain that even the order of these events will not be preserved from performance to performance. The 7th unit in a 13-in-the-time-of-1 grouping, and the 6th unit in an 11-in-the-time-of-one grouping, over the same crotchet, at [crotchet = 120] are only 1/143th of a crotchet, or 1/286th of a second (less than 4 milliseconds) apart. It only needs one of these units to be misplaced by 3 or more milliseconds in one direction, and the other by 3 or more milliseconds in the opposite direction, for the order of the two events to be reversed in live performance! (See **Diagram 12**).

Hence we are not, here, composing a sequence intrinsically tied to an "exact" notation. The exact notation in fact specifies a class of sequences with similar properties within a fairly definable range. We are in some senses specifying a density flow within the limitations of a certain range of random fluctuations. We could describe this class of sequences by a time-varying density function with a specified (possibly variable) randomisation of relative time-positions. In writing notes on paper, the exact notation is more practicable so long as we acknowledge that it does not specify an exact result. In the computer domain, the density approach may be more appropriate if we are really looking for the same class of *percepts* as in the notated case.

In such complex cases, we may still retain a longer time-frame reference set (see **Diagram 13**).

In a cyclically repeated pattern of superimposed "irrationally" related groupings, even given the intrinsic "inaccuracy" of live performance, we should be aware of the repetition of the bar length or larger-time-frame unit (which we are dividing). We hear in a measured way at the larger time-frame level, but we hear only comparatively at smaller time-frames. On the other hand, with computer-precision in generating the sound sequence, we may be able to perceive a very precise "density flux" in the combination of these streams, at least if they are repeated a sufficient number of times. **(Sound example 9.9)**.

Our comments about notated music are further reinforced if we now organise the material internally so as to contradict any mutual reinforcement at larger pulse (e.g. bar) interfaces. In this way we can also destroy measured perception at the larger level. Again, if we repeat a sequence of (say) 4 bars, we may re-establish a measured perception of phrase regularity in a yet larger time-frame, e.g. the 4-bar frame. **(Diagram 14)**.

Eventually, however, either because we change bar-lengths in a non-repeating way, or because we constantly undermine the bar-level mutual pattern reinforcement, larger time-frame reference-frames will not be perceptually established. **(Diagram 15)**.

We then pass over exclusively to comparative or to textural perception.

DIAGRAM 13

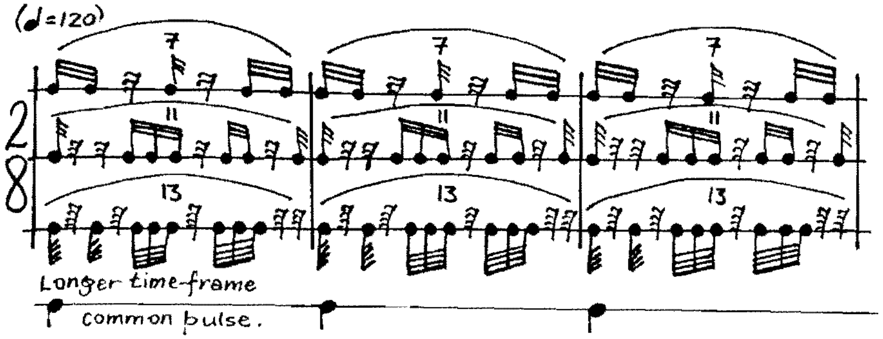


DIAGRAM 14

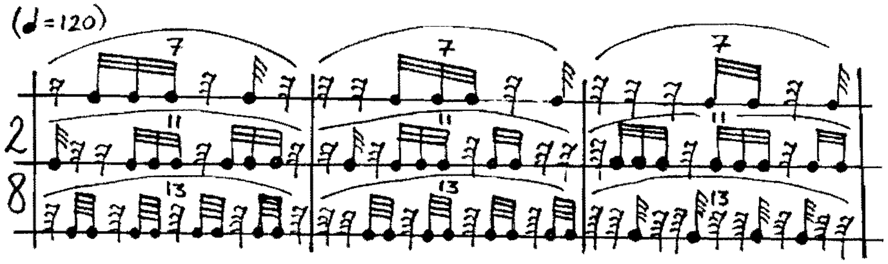
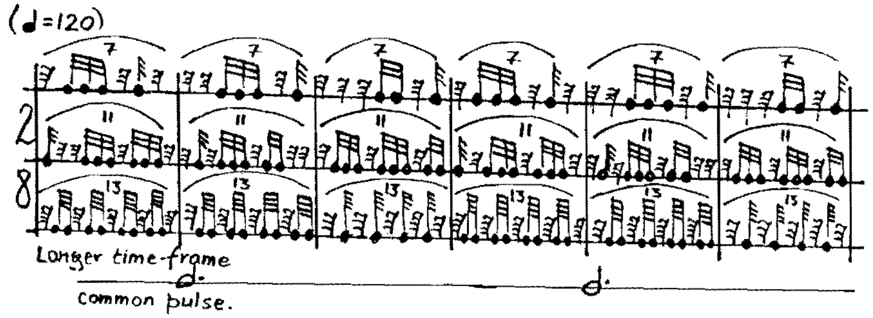


DIAGRAM 15



LARGE-DURATION REFERENCE FRAMES - THE NOT SO GOLDEN SECTION

At the other extreme, music can be constructed as a set of nested time-frames in which at one level (e.g. a semiquaver time-frame) order-sequences are put in motion (rhythm) while at the same time establishing a time-frame (e.g. semibreve bars) in which longer duration order sequences can be set up.

Within the time-frames fast-demisemiquaver to 1-or-2-minutes, hierarchical systems of interlocked levels functioning as order sequences in one direction and duration frames in the other can be articulated. **(Diagram 16).**

There will, however, be a limit to our ability to perceive a large time-frame as precisely measurable and hence comparable with another event of comparable time-frame proportions, i.e. a timescale limit to measured perception. Though many composers writing musical scores would stake their reputations on the idea that we can hear and respond to precise proportions in larger time-frames (especially those derived from the Fibonacci series and associated Golden Section) there is little evidence to support this. Such proportions certainly *look* nice in scores and provide convenient ways to subdivide larger time-scales.

The exact proportions offered by the Fibonacci series have the particular advantage of being nestable at longer and longer time-frames. **(Diagram 17).**

This makes them attractive for work based on integral time units, e.g. quavers at the fundamental tempo of a piece. The question is, however, can we perceive these proportions in a measured sense, or do we perceive them in a comparative sense (1 to 1/2ish-2/3ish) while the exact measure makes the task of laying out the *score* possible. If we consider the golden section itself (the limit of the Fibonacci series) this will nest perfectly with itself. **(Diagram 18).**

However, being an irrational quantity (in the mathematical sense - it cannot be expressed as the ratio of *any* pair of whole numbers) its parts are not integrally divisible by any number of fixed pulses, no matter how small. Hence it is intrinsically problematic to attempt to score in exact Golden Sections. There is, of course, no such intrinsic problem to *cutting* tape-segments to such irrational proportions and hence, hypothetically, a perfect Golden Section nesting might be achieved in a studio composition. Whether we are *measurably* perceiving this handiwork is quite a different question.

On the relatively small time-frame of "swing" (see earlier discussion), subtle but *comparative* perception of proportions may be very significant (just as there are very subtle parameters in the spectral character of grain which we cannot "hear out" but which are fundamental to our qualitative perception). However, on much larger time-frames, we run into the problem of both longer term time memory and the influence of smaller scale event patterns on our sense of the passage of time (waiting ten minutes for this train seemed an interminable time .... reading the newspaper, an hour has gone by without my noticing).

Due to the structure of the Fibonacci series, almost *any* proportion between 1 to 1/2 & 1 to 2/3 can be found within it and the nearer these proportions are to the Golden Section, the more likely they are to occur in the series. The question is, do we believe that we can measurably perceive these proportions (e.g. 34:21) or merely that we can comparatively perceive them all as lying within the range 1 to 1/2ish-2/3ish. More dramatically, because the Golden Section is an irrational quantity, then, by

DIAGRAM 16

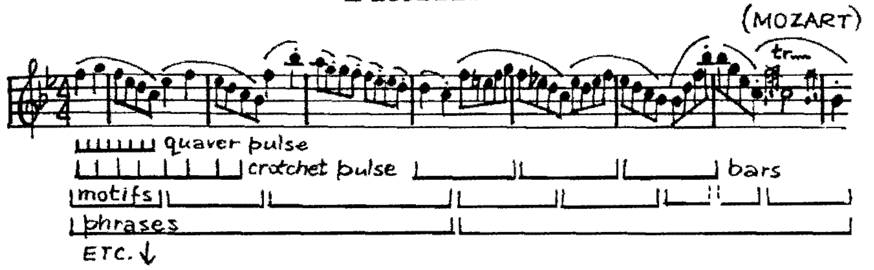
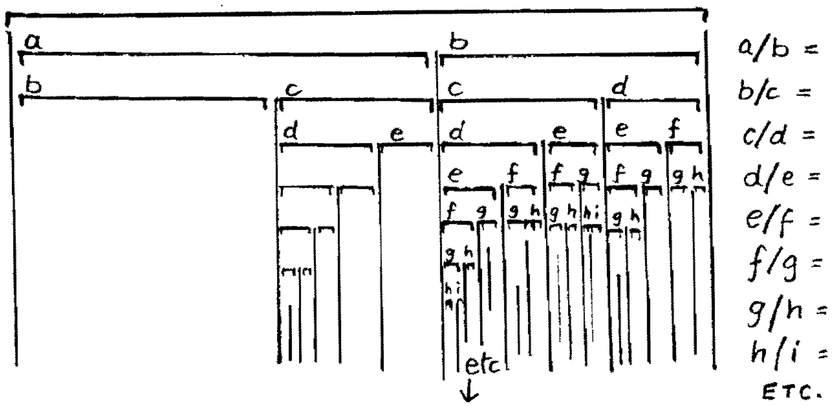


DIAGRAM 17

DIAGRAM 18



definition, there is no smaller time-frame over which it can be measurably perceived. This is part of the definition of a mathematically irrational quantity. Hence, by definition, we cannot measurably perceive a Golden Section, though again we may comparatively perceive that it lies within the bounds of 1 to 1/2ish–2/3ish.

I am happy to accept that Fibonacci derived ratios give a satisfactory comparative percept of 1 to 1/2ish–2/3ish (possibly even a slightly narrower range) but not that, at larger time-frames, there is any measured perception involved. In large time-frames it is certainly not possible to distinguish a Golden Section from a host of other roughly similar proportions lying within a range. The longer the time-frame the larger this range of inexactness becomes.

I am certain of two things. Firstly that when psycho-acousticians run detailed tests, even on composers who regularly use these devices, they will discover that these relations on larger time-frames are not heard exactly. Secondly, that when this is discovered, a large section of the musical community will reject the results. Number mysticism has a long and distinguished tradition in musical thought and the Golden Section is a well-defended tenet of a modern musical gnosticism.

WHEN IS DENSITY CONTROL APPROPRIATE ?

We must now ask the crucial question, when is composing with density parameters appropriate and when, conversely, should we proceed in a deterministic way, i.e. specifying exact time-placement, exact pitch-value, exact spectral contour etc.

If we demand computer precision, we will produce a precise, and precisely repeatable, result. The question here is, how exact does it *need* to be, i.e. by how much can we alter the parameters before we notice any difference in what we perceive. This judgement has to be set in context. Is the musical organisation focusing our attention on these exact timing aspects of the total gestalt, e.g. through cyclical repetition of the pattern in which timing parameters are slowly changing in a systematic manner? Or is the time-sequence within this event merely one of several parameters the listener is being asked to follow? Or is it an incidental result of other processes and not central to the way in which we scan the musical events for a sense of mutual connectedness and contrast?

It is always possible to *claim* that everything is important. In particular, composers are especially prone to claiming that anything is an important constituent of what the listener hears if it was an important part of the *method* by which they arrived at the sounding result. However, this need not be the case. I can devise a method for producing sound in which every wavecycle has at least one sample of value 327. This, *by itself*, will provide no perceived coherence within the sound world I create. To declare that it does is simply dishonest. The composer must thus be able to monitor what is and what is not perceptibly related in a composition, independent of the knowledge he/she has of the generating procedure.

Deterministic procedures in which parameters are systematically varied may be used to generate ranges of sound materials. We need, however, a separate approach, based in perception, to define whether the sound materials generated *are* related and in what ways they are related, if at all. The generating procedure and its systematicness are not, on their own, a guarantee of perceptual relatedness.

We may, in the same spirit, use *texture generation* procedures, also with systematic variation of parameters, to generate materials for subsequent perceptual assessment.

Conversely, through experience, we may already know what kind of materials we want. If, for example, we know that the result must be characterised by slowly density-fluctuating arhythmicity, it would be more direct and simpler, when composing sound directly with a computer, to use a *texture control* process, refining its parameters until we get the result we desire. Defining a deterministic process to attain the same perceptual result will usually be much more difficult both to specify and to systematically vary in some way *coherent with perceptual variations.*

We may also combine the two approaches by employing *stochastic processes* in which the probabilities of proceeding from one set of states to the next are defined precisely, ensuring a high degree of control at one time-frame over a process that evolves unpredictably at smaller time-frames. As the transition probabilities can have any value up to unity (completely deterministic) we have here a way to pass seamlessly between pure density perception and total rhythmic determinism (or between the equivalent extremes for any other sonic parameter). Again we must be clear that whether the result is rhythmic or arhythmic, density-perceived or field-perceived, is to be ascertained by listening to the results, not by appealing to the method of generation.

Compositional virtue does not lie in the determinism (or even the describability) of the compositional method, but in the control of the perceived results and their perceptual connectedness. This is unfortunately not always the view handed down at academies of music composition.

CHAPTER 10

ENERGY

ENERGY TRAJECTORIES

Sounds in the real world which do not simply die away to nothing require some continuous energy input such as bowing, blowing, scraping or an electrical power supply to maintain them. Usually the level of the flow of energy into the system determines the loudness of the output and hence allows us to read the fluctuating causality of the sound.

We may create the illusion of energy flow with a continuous, grain-stream, sequential or textural sound by simply imposing a loudness trajectory, or envelope, on it (*enveloping*) (**Sound example 10.1**).

Such a trajectory may be derived directly from the activity of another sound (prerecorded or a live performer) (*envelope following*). Enveloping is particularly useful for creating a loudness anacrusis (a crescendoing onto a key sonic event), a particularly musical device as most natural sounds have precisely the opposite loudness evolution (**Sound example 10.2**).

Conversely, we may begin with a sustained sound (continuum, grain-stream, sequence, texture) and give it an exponentially decaying loudness, through *enveloping*. This suggests the sound originates from some vibrating medium which has been struck (or otherwise set in motion) and then left to resonate, especially if the onset of the sound is reinforced. With simple sustained sound this can completely alter the apparent physicality and causality of the source (see below). With sounds which more easily betray their origins (e.g. speech) we produce an interesting dual percept. (**Sound example 10.3**).

If a sound already has a noticeably varying loudness trajectory we may track this variation (*envelope following*) and then perform various operations with or on this trajectory (*envelope transformation*). As already suggested we might transfer the trajectory to an entirely different sound (*enveloping* or *envelope substitution*). Sequences of events which are extremely strongly characterised by loudness articulation may thus be 'recapitulated' with an entirely different sonic substance (**Sound example 10.4**).

We may also modify the loudness trajectory we have extracted (*envelope transformation*) and reapply it to the original sound. Note that we must do this by *envelope substitution* and not simply by *enveloping* (see **Appendix pp59 & 61**). Thus we may exaggerate the loudness trajectory to heighten its energetic (or dramatic) evolution (*expanding* : see below) .

The loudness trajectory may also be lengthened (*envelope extension*) or shortened (*envelope contraction*), extending or contracting the associated gesture (see for example the discussion of *time-stretching* textures in Chapter 11), and any of the trajectory (envelope) transformations described here might be applied in a time-varying manner so that e.g. a sound may gradually become prominently *corrugated* (see below).

FINE-GRAINED ENVELOPE FOLLOWING

The result of the process of envelope following will depend to a great extent on how the loudness trajectory is analysed. To assess the loudness of a sound at a particular moment, we need to look at a certain small time-snapshot of the sound (a time-frame defining a window size : see **Appendix p58**). The *instantaneous* loudness of a sound has no meaning (see below). The size of this time-snapshot will affect what exactly we are reading.

Thus if we wish to exaggerate the loudness trajectory of a rhetorical speech, we are interested in detecting and exaggerating the variations in loudness from word to word or even from phrase to phrase. We do not wish to exaggerate the loudness variation from syllable to syllable. Even less do we wish to exaggerate the undulations of loudness within a single rolled 'r'. We want a "coarse" reading of the loudness trajectory. We therefore choose a relatively large time-frame over which the loudness of the signal is measured. (Alternatively we might detect the loudness level at the smallest meaningful time-frame (around 0.005 seconds) and then average the result over an appropriate number of these frames to get a coarser result).

Conversely, we may wish to track loudness changes more precisely. If loudness trajectory is tracked using a fine time-frame (small window) over a granular sound, we will often be able to detect the individual grains. Exaggerating the fluctuations in the trajectory can then be used to force momentary silences between the grains (*corrugation*). This both alters the quality of the sound and makes the grains more amenable to independent manipulation (**Sound example 10.5**).

Following or manipulating the loudness trajectory below the grain time-frame has little meaning. The instantaneous value (pressure) of the sound wave changes from positive (above the norm) to negative (below the norm) to create the experience we know as sound. The amplitude of the signal describes the size of these swings in pressure, and, as we cannot know how large a swing will be until it reaches its maximum excursion, we need to use a time- window which covers whole wavecycles when determining the loudness trajectory of a sound.

Beneath the duration of the wavecycle the instantaneous variation of the pressure determines the shape of the wave and hence the sound spectrum. Imposing loudness changes over one (or only a few) wavecycles, or wavesets, in effect changes the local shape of the wave(s) and is hence a process of spectral manipulation rather than of loudness control (see **Appendix p53**). But if we gradually expand the number of wavecycles or wavesets we cause to fall under our loudness-changing envelope, we pass from the spectral domain to the time-frame of grain-structure and eventually to that of independent events. These transitions are nicely illustrated by *waveset enveloping*. In **Sound example 10.6**, the number of wavesets falling under a single loudness-changing envelope is progressively increased on each repetition.

The foregoing discussion illustrates once again the importance of time-frames in the perception and composition of sonic events.

GATES AND TRIGGERS

We may apply more radical modifications to the loudness trajectory of a sound. We may cause a signal to cut out completely if it falls below a certain level. This procedure, known as *gating*, can be used for elementary noise reduction, but is also often used in popular music to enhance the impact of percussive sounds.

Detecting the appropriate level to apply the gate is known as threshold detection. We may also use threshold detection as a means to *trigger* other events, either when the level of a signal exceeds a threshold, or when it falls below it. In the former case we may heighten the attack characteristics of particularly loud sounds by causing them to trigger a second sound to be *mixed* into the sonic stream. Or we may trigger quite different sounds to appear briefly and at low levels, so that these new sounds are partly masked by the original sound, making a kind of hidden appearance only (**Sound example 10.7**).

Alternatively we might trigger a process affecting the source sound, such as *reverberation* or *delay*. We could switch such processes on and off rapidly, or vary their values according to the loudness trajectory of the source signal, e.g. quiet events might be made more strongly reverberant and louder events drier. Combining this with triggered spatial position control we might articulate a whole stereo-versus-reverb-depth space from the loudness trajectory of a single mono source.

In conventional recording studio practice two very commonly used gating/triggering procedures are *limiting* and *compressing*. In a limiter, any sound above a certain threshold loudness is reduced in level to that threshold value. This ensures that e.g. a concert can be recorded at a high level with less risk of particularly loud peaks overloading the system. Compression works in a more sophisticated manner. Above the threshold the louder the sound, the more its loudness is reduced, producing a more subtle containment of the signal. This process may be used more generally, setting the threshold level quite low, to flatten or smooth the gestural trajectory of a sound, thereby e.g. "calming" a hyperactive sequence of events.

An *expander*, in contrast, does the opposite, expanding the level of a sound more, the louder it is, hence exaggerating the contrasts in loudness in a sound source and thereby perhaps exaggerating the gestural energy of a sound event or sequence of events. (See **Appendix p60**).

BALANCE

In orchestral music, the balance of loudness between sound sources is achieved partly through the combination of performance practice with the notation of dynamics in a score, and partly through the medium of the conductor who must interpret the composer's instructions for the acoustic of the particular performance space. Furthermore, the blending or contrast of sounds is aided or hindered by the fact that all the sounds are generated in the same acoustic space (normally). In the studio we may bring together sounds from entirely different acoustic spaces (a forest, a living room) and with quite different proximity characteristics (close-miked, or miked at a distance such that room ambience is significantly incorporated into the sound).

Moreover, we may combine such features in ways which contradict our experience of natural acoustic environments. Thus proximately recorded loud sounds may be played back very quietly, while distant quiet sounds may be projected with very loudly. These features of the sound landscape are discussed more fully in *On Sonic Art*.

In sound composition in general, the loudness balance between diverse sources is described in a (possibly graphic) *mixing score*, or created in real-time at a mixing desk, perhaps with action information recorded for subsequent exact reproduction or detailed modification.

This can give us very precise control over balance, even within the course of grain size events. Balance changes within the grain time-frame is in fact a means of generating new sonic substances intermediate between the constituents (e.g. *inbetweening*, see Chapter 12). (**Sound example 10.8**).

Mixing may also be used to consciously mask the features of one sound by another or, more usually, to consciously avoid this. The latter process is aided by distributing the different sounds over the stereo space. (**Sound example 10.9**).

In certain cases we may wish to ensure the prominence of a particular sound or sounds without thereby sacrificing the loudness level of other sources. In popular music the device of *ducking* is used to this end. Here the general level of some of the other instruments is linked inversely to that of the voice. Before the singer begins, the rest of the band is recorded as loudly as possible, but, once the voice enters, some instruments 'duck' in level so as not to mask the voice. This ensures that the voice is always clearly audible while at the same time a generally high recording level is maintained whether or not the voice is present.

This process may be used more generally. We may *mix* any two sounds so that the loudness trajectory of the second is the inverse of the first (*envelope inversion*). For example, a sound of rapidly varying loudness (like speech) may be made to interrupt a sound of more constant energy (a large crowd talking amongst itself, heavy traffic) by imposing the inverted loudness trajectory of the speech on the traffic noise (**Sound example 10.10**). This procedure will ensure that the traffic will not mask the voice, yet will remain prominent, perhaps alternating in perceptual importance with the voice (depending on how the relative maximum levels are set and on the intrinsic interest of the traffic sounds themselves!).

PROXIMITY

Loudness information in the real world usually provides us with information about the proximity of a sound source. The same sound heard more quietly is usually further away. It is important to understand that other factors feature in proximity perception, especially the presence or absence of high frequencies in the spectrum (such high frequencies tend to be lost as sound travels over greater distances). There are of course other aspects of the sound environment affecting proximity perception. The presence or absence of barriers (walls, doors to rooms or entrances to other containers) and the nature of reflective surfaces (hard stone, soft furnishings) contribute to a sense of ambient *reverberation*. Even air temperature is important (on cold clear nights, sound waves are refracted downwards and hence sounds travel further).

Using a combination of loudness reduction and *low pass filtering* (to eliminate higher frequencies) we may make loud and closely recorded sounds appear distant, especially when contrasted with the original source. But we may also produce self-contradictory images e.g. closely miked whispering, which we associate with close-to-the-ear intimacy and hence "quietness", may be projected very loudly, while bellowed commands may be given the acoustic of a small wooden box – and these may both be projected in the same space.

Proximity may also be used dynamically, like a continuous zoom in cinematography. A microphone may be physically moved towards or away from a sound source during the course of a recording. The effect is particularly noticeable with complex spectra with lots of high frequency components (bells and other ringing metal sounds) where the approach of the microphone to the vibrating object picks out more and more high frequency detail. This aural zoom is hence a combination of loudness and spectral variation. It is important to understand that this effect takes place over a very small physical range (a few centimetres) in contrast to the typical range of the visual zoom. (**Sound example 10.11**).

PHYSICALITY AND CAUSALITY

As we have already suggested, loudness trajectory plays an important part in our attribution of both the physicality of a source (rigid, soft, loose aggregate etc.) and the causality of the excitation (struck, stroked etc). The loudness trajectory of the sound onset is particularly significant in this respect, and often simple onset-trajectory manipulation is sufficient to radically alter the perceived physicality/causality of the source. Thus almost any sound can be given a struck-like attack by providing a sudden onset and perhaps an exponential decay. Conversely, a strongly percussive sound may be softened by very carefully 'shaving off' the attack to give a slightly more slowly rising trajectory. (**Sound example 10.12**).

This area is discussed in more detail in Chapter 4.

CHAPTER 11

TIME STRETCHING

TIME-STRETCHING & TIME-FRAMES

Time-stretching and time-shrinking warrant a separate chapter in this book because they are procedures which may breech the boundaries between perceptual time-frames. The importance of time-frames in our perception of sounds is discussed in detail in the section "Time-frames: samples, wavecycles, grains & continuation" in Chapter 1.

Furthermore, the degree of time-stretching of a sound may itself vary through time, and with the precision of the computer such time-varying time-stretching (*time-warping*) may be applied with great accuracy.

There are several different approaches to time-stretching and the approach we choose will depend both on the nature of the sound source and the perceptual result we desire. Below we will look at *tape-speed variation*, *brassage* techniques, waveset repetition (*waveset time-stretching*), frequency domain time-stretching (*spectral time-stretching*), and grain separation or grain duplication (*granular time-stretching*). We will then discuss various general aesthetic and perceptual issues relating to time- stretching before dealing with the most complex situation, the time-stretching of texture streams.

"TAPE-SPEED" VARIATION

In the classical tape-music studio, the only generally available way to time-stretch a sound was to change the speed at which the tape passed over the heads. The digital equivalent of this is to change the sampling frequency (in fact, interpolating new sample values amongst the existing ones, but storing the result at the standard sampling rate : see **Appendix p37**). This approach is used in sampling keyboards (1993) and table-reading instruments e.g. in *CSound*.

In both cases the procedure (*tape-speed variation*) not only changes the sound duration but also the pitch because it alters the wavelengths (and therefore the frequency : see **Appendix p4**) in the time-domain signal. Similarly, it changes the frequency of the partials and hence also shifts the spectral contour, and hence the formants. And it time-stretches the onset characteristics, probably radically changing the sound percept in another way. (**Appendix p36**). (**Sound example 11.1**).

Although time-warping, pitch-warping and formant-warping are thus not independent, this approach has its musical applications. In particular (multi-) octave (etc) upward transpositions can be used as short time-frame reinforcements of a sound's onset characteristics (see Chapter 4). Moreover, downward transposition by one or two octaves not only reveals the details of a sound's evolving morphology in a slower time-frame, making important details graspable. The transposition process itself often brings complex, very high frequency spectral information into a more perceptually accessible mid-frequency range (our hearing is most sensitive in this range). The internal qualities of a complex sound may thus be magnified in two complementary domains. Digital recording and transposition is

also much more robust than the old analogue recording tape-speed variation, capturing important extremely high frequency information for re-listening in a moderate frequency range and preserving frequency information transposed down to very low frequencies. (**Sound example 11.2**).

Tape-speed variation may be applied in a time-varying manner e.g. causing a sound to plunge into the lowest pitch range, hence bringing very high frequency detail into the most sensitive hearing range, at the same time as magnifying the time-frame. Conversely, a sound may accelerate rapidly, rising pitch-wise into the stratosphere (using e.g. *tape acceleration*). As the sound rises, internal detail is lost. With sufficient acceleration almost any sound can be converted into a structureless rising pitch-portamento. This is an elementary way to perceptually link the most diverse sound materials, if they occur in long enough streams for such acceleration to be possible, or a way of creating musical continuity between a complex stream of diverse events and event structures focused on pitch portamenti. (**Sound example 11.3**).

BRASSAGE TECHNIQUES

As discussed previously, *Brassage* involves *cutting* a sound into successive, and possibly overlapping, segments and then reassembling these by *resplicing* them together (for pure time-stretching, in exactly the same order) but differently spaced in time (see **Appendix p44-AB**) . It can always be arranged, by appropriate choice of segment length and segment overlap, for the resulting sound to be continuous, if the source sound is also continuous. Provided the cut segments are of short grain duration (i.e. with perceptible pitch and spectral properties but no pitch or spectral evolution over time) then the goal sound will appear time-stretched relative to the source.

Good algorithms for doing this are currently (1994) embedded in commercially available hardware devices (often known as *harmonisers*) and function reasonably well, often in a continuously time-variable fashion over a range half to two-times time-stretch. At the limits of this range and beyond we are beginning to hear spectral and other artefacts of the process. These may, however, be useful as sound-transformation techniques. (**Sound example 11.4**).

Particularly in long time-stretches, Brassage may lead to...

(1) pitch artefacts – related to the event-separation rate of the segments.

(2) granulation artefacts – where the individual grains are large enough to reveal a time-evolving structure, and hence, as successive segments are chosen from overlapping regions of the source, delayed repetitions are heard.

(3) phasing artefacts – due to the interaction of rapid repetition or "delay", and gradual shifting along the source.

With long time-streches the perceptual connection between source sound and goal sound may be remote and may require the perception of mediating sounds (with less time-stretching) to make the connection apparent. Repeated application of Brassage techniques to a source (in effect using "Feedback") may entirely destroy the original characteristics of the source. In contrast, *spectral time-stretching* (in the frequency domain) can be repeated non-destructively. (**Sound example 11.5**).

Brassage may be extended in a variety of ways, using larger and variable length segments, varying the time-range in the source from which the goal segment may be selected, varying the pitch, loudness and/or spatial position of successive segments and, ultimately, varying the output event density. We thus move gradually out of the field of time-stretching into that of generalised brassage and granular reconstruction. (These possibilities are discussed in more detail in Chapter 5 in the section "Constructed Continuation" and in **Appendix pp44-45** and **Appendix p73**). We need add only that using Granular Reconstruction with time-varying parameters (average segment length, length spread, search range, pitch, loudness, spatial divergence and output density) it is possible to create, in a single event, a version of a source sound which begins as mere time-stretch and ends as a texture stream development of that source.

A more time-stretching satisfactory application of the Brassage process to sequences such as speech streams can be achieved by source-segment-synchronous brassage. In this case we need to apply a sophisticated combination of *envelope following*, and pitch-synchronous spectral analysis to isolate the individual segments of the source stream. These can then be individually brassage-time-stretched and *respliced* together in one operation. Because none of the goal segments crossover between source segments, we avoid artefacts created at source segment boundaries in simple brassage (see **Diagram 1**).

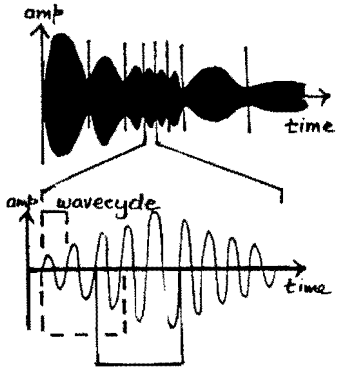
WAVESET TIME-STRETCHING

Time-stretching can be achieved by searching for zero-crossing pairs and repeating the wavesets thus found (see **Appendix p55**). This technique will produce only integral time-multiples of the source duration. As discussed elsewhere, two zero-crossings do not necessarily correspond to a wavecycle (a true wavelength of the signal) so a waveset is not necessarily a wavecycle. As a result this technique will have some unpredictable, though often interesting, sonic consequences. Sometimes parts of the signal will pitch-shift (e.g. at x2 time-stretch, by an octave downwards, as in *tape-speed variation*). For a x2 (or even x3) time-stretch, artefacts can often be reduced by repeating pairs (or larger groups) of wavesets (a special case of pitch- synchronous brassage). These kinds of artefacts can of course be avoided, in truly pitched material, by using a pitch-following instrument to help us to distinguish the true wavecycles. (**Sound example 11.6**).

As the number of repetitions increases, other artefacts begin to appear. At x3 there is often a "phasing"-like coloration of the sound. With time evolving and noisy signals, at x16 time stretch a rapid stream of pitched beads is produced as each waveset group achieves (near-) grain dimensions and is heard out in pitch and spectral terms. (No such change occurs, however, in a steady tone or stable spectrum.) This spectral fission is heard "subliminally" within the source even at x4 time-stretch. (**Appendix p55**). (**Sound example 11.7**).

It is possible also to interpolate between waveset durations and between waveset shapes through the sequence of repetitions. In x64 time-stretching with such interpolation the new signal clearly glides around in pitch as each "bead" pitch glides into the next. At x16 time-stretch, we are aware more of the "fluidity" of a bead-stream rather than of a continually portamentoing line. Even at x4 time-stretch, this fluidity quality contributes to the percept in an intangible way. (**Sound example 11.8**).

DIAGRAM 1

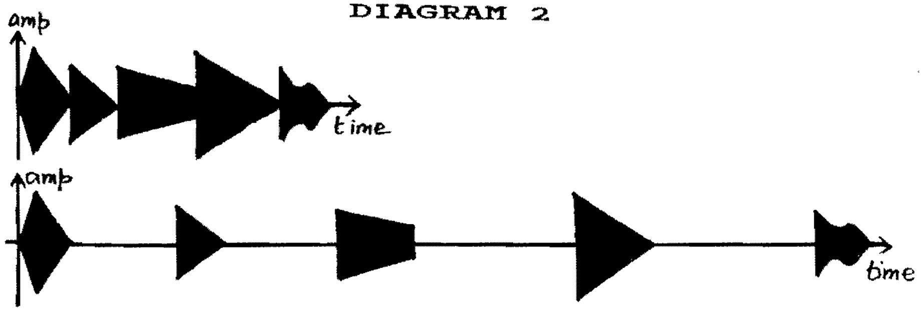


Use fine-grained envelope following and detailed spectral analysis to extract sequential features of the sound.

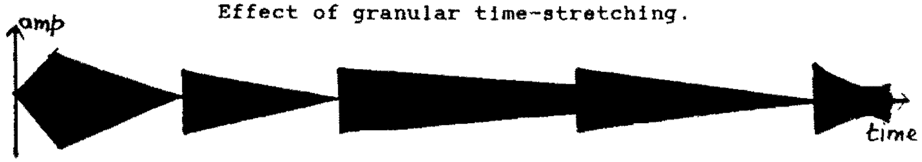
Deduce pitch of feature.

Use pitch to help define pitch-synchronous brassage segments of feature.

DIAGRAM 2



Effect of granular time-stretching.



Effect of spectral time-stretching.

Except for x2 time-stretching, when working on musically interesting sound sources, waveset repetition is more useful as a special sound transformation procedure. Successive applications of such x2 stretching does not significantly reduce the bead-streaming effect in long time dilations.

SPECTRAL TIME-STRETCHING

A sustained sound with stable spectrum and no distinctive onset characteristics may be analysed to produce a (windowed) frequency-domain representation. We can then resynthesize the sound, using each window to generate a longer duration (*phase vocoder*). The resultant sound appears longer but retains its pitch and spectral characteristics. We may notice an extension of the initial rise time and final decay time but in this case these may well not be perceptually crucial. Of all the techniques so far discussed, this is by far the best for pure time-stretching and works well up to about x8 time stretching. (**Sound example 11.9**).

Beyond this,however, it too becomes less satisfactory as a pure time-stretching procedure. The original window size for the analysis is chosen to be in the time-domain so brief that the human ear perceives the small change from window to window (in fact a "step") as a smooth continuous transition. Once we do too long a resynthesis from individual windows, e.g. a x64 time-stretch, the spectrum of each window is sustained long enough for us to become aware of the joins. The continuity of the original source is not reconstructed.

One way around this limitation is to stretch the source x8, synthesize the result, reanalyse and time-stretch by x8 once again. However, a more satisfactory approach is to interpolate between the existing windows to create new windows intermediate in channel-frequency and channel-loudness values between the original windows but spaced at the original window time-interval. (*spectral time-stretching*). The procedure ensures a perceptually continuous result even at x64 time stretch. (**Sound example 11.10**).

However, even in this case, spectral transformations may arise. In particular, a sound with a rapidly changing spectrum, originally perceived as noise, will be sufficiently slowed for us to hear out the spectral motion involved. In general we will hear a resulting sound with a gliding inharmonic ("metallic") spectrum in place of our source noise, at very great time- stretching factors. (**Sound example 11.11**).

Using time-variable spectral time-stretching (*spectral time-warping*), we can use this effect to produce spectrally diverse variants of a source, zooming in to a maximal time-stretch at a particular point in a source, will produce a particular inharmonic artefact in the goal sound. Zooming in to a different point in the source will produce a different inharmonic artefact in the new goal sound. We can thus produce a collection of related musical events deriving from the same source (provided enough of the resulting signals are elsewhere similar to each other). (**Sound example 11.12**).

The ultimate extension of this process is *spectral freezing*, where the frequencies or loudnesses of the spectral components in a particular window are retained through the ensuing windows. This compositional tool is discussed in Chapter 3.

If we have a sound with marked (grain time-frame) onset characteristics, e.g. an attack-dispersion sound (piano, bell, pizz), even our interpolated *spectral time-stretch* over a long stretch, will radically

alter the sound percept by making the indivisible qualitative character of the onset become a time-varying percept (see Chapter 1). (**Sound example 11.13**).

If we wish to preserve the characteristics of the source sound, we must retain the onset characteristics by not time-stretching the onset. This we can achieve by time-variable spectral time-stretching (*spectral time-warping*), making the time-stretch equal to 1.0 (i.e. no stretch) over the first few milliseconds of the source and thence increasing as rapidly as we wish to any large value we desire. (**Sound example 11.14**).

Alternatively, we can use spectral time-stretching to heighten the internal spectral variation of the source. Time-stretching the onset of the signal, but not the continuation, will alter the sound percept radically (altering the causality) but retain a perceptual connection between source and goal through the stability of the continuation. Clearly, the longer the continuation, the stronger the sense of relatedness. Clearly also, there are many combinations of onset transformation and continuation transformation. (**Sound example 11.15**).

In a very long spectral time-stretch of a sound's continuation, where the sound is spectrally varying, we can reveal these changes by stripping away partials from the time-stretched sound until only the most prominent (say) twenty remain (*spectral tracing* : see **Appendix p25** and Chapter 3). As the spectrum varies in time, partials will enter and leave this honoured set and we will hear out the new entries as 'revealed melodies', a type of *constructive distortion*. If the time-stretch is itself not time-varying, these revealed melodies will be regularly pulsed at the time-separation of the original window duration multiplied by the time-stretch factor. *Spectral time-warping* will create a fluid (accel-rit) tempo of entries and will accelerate into the time continuum where the time-stretch factor reduces towards 1.0 (no stretch). (**Sound example 11.16**).

GRANULAR TIME-STRETCHING

The time-stretching of grain-streams is problematic. As we have seen, if we time-stretch the onset of a sound we risk completely altering its perceived character. We overcame this problem by time-variably time-stretching (*time-warping*) a sound in such a way that the onset was not stretched. However, grain-streams are in effect a sequence of onsets. We cannot in this case, therefore, preserve only the beginning of the sound. Ideally we would need to use an *envelope-follower* to uncover the loudness trajectory of the sound and thus locate all the onsets, and then apply a time-warping process that left the sound unaltered during every onset moment. This is feasible but awkward to achieve successfully.

It is therefore useful to be able to time-vary a grain-stream by separating the grains and repositioning them in time, causing the sequence of grains to accelerate, ritardando, randomly scatter etc. (*granular time-warping by grain separation*). (**Sound example 11.17**).

With even moderately large granular time-stretching of this sort, the stream character of a grain-stream breaks down in our perception - we hear only isolated staccato events, the elements of potential musical phrases. Conversely, time-shrinking of a sequence of isolated events, by reducing separation time, can reach a point where the sounds become a grain-stream, or sequence-stream, rather than musical "points" in their own right. (**Sound example 11.18**).

We may also extend grain-streams by duplicating elements, (*granular time-stretching by grain duplication*) e.g. respace the original units at twice the initial separation and repeat each grain halfway in time between the new grains. This is a grain dimension analogue of waveset repetition and suffers from the same drawbacks. If the grain elements are changing in character, grain repetition will be perceptually obvious and the more repetitions of each grain, the more perceptually prominent it will be. (It might be obviated by a very narrow range pitch-scattering of the grains to destroy the sense of patterning introduced by the grain repetition) Note that in this way we can time-stretch the grain-stream by integral multiples of the original duration (and by selected choice of grains to repeat, by any rational fraction), without thereby altering the (average) pulse-rate (density) of the stream. **(Sound example 11.19).**

TIME-SHRINKING

All the above processes may be applied to time-contraction, with noticeably different results. *Tape-speed variation* time-contraction has already been discussed. *Spectral time-shrinking* can smoothly contract any sound, or part of a sound. It can be used to contract a sound with continuation into an indivisible grain, though, in this process data will be irretrievably lost, i.e. if the grain is now time-dilated once more, the resulting sound will be considerably less time-variably spectrally detailed than the original. The process of successive contraction and expansion will give similar results to the process of spectral blurring (see Chapter 3). **(Sound example 11.20).**

If grain-streams or sequences are *spectrally time-shrunk*, the individual elements will shrink, become less spectrally detailed and more click-like as their durations approach the lower grain time-frame boundary. **(Sound example 11.21).** *Granular time-shrinking by grain separation* involves the repositioning of grains by reducing intervening silence (where possible), or overlaying or *splicing* together the existing grains. These latter processes will tend to blur grains together. The process hence tends towards tremolo-like loudness trajectory and eventually to a continuous percept, from which the original grain is not recoverable. (A new grain might be created by *enveloping*.) **(Sound example 11.22).**

Granular time-shrinking by grain deletion is not prone to this blurring effect, but it is easy to destroy the continuity of the grain-stream percept if grains are deleted from that stream.

In *waveset time-shrinking* waveset duplication is replaced by waveset omission. This process also loses granular or sequential detail as the sound contracts, though in a different way to grain separation contraction. We may choose to omit wavesets in a sequentially regular manner (e.g. every fourth waveset) in which case the sound becomes increasingly fuzzy, quieter, less detailed. Or, we may choose to omit the least significant (lowest amplitude) wavesets. In this case the sound retains its original loudness and loudness trajectory, but also loses sequential detail. The first process tends towards silence, the latter eventually reduces any sound to a continuation-less but loud point sound. **(Sound example 11.23).**

THE CONSEQUENCES OF TIME-WARPING : TIME-FRAMES

As we have discovered time-warping is not a single, simple process. The musical implications of applying particular processes to particular sounds must be considered on their own merits. There are, however, a number of general perceptual considerations to be borne in mind.

Considering single grain time-frame sounds, the internal structure of the grain, perceived as in indivisible whole, a quality of the event, when time-stretched becomes a clearly time-varying property, a feature of the sound continuations . We may in this way uncover changing pitch, changing spectral type, changing formant placement, changing loudness undulations in any of these, or specific granularity of sequencing of events within the original grain. In many cases there may be no apparent perceptual connection between the quality of the grain and the revealed morphology of a much time-stretched version of that grain. It will need to be established through the structural mediation of sounds made by time-stretching the grain by much less. In fact, perceptual connectedness will be found to follow an exponential curve as we move away from the grain dimensions i.e. initially very tiny spectral time-stretches will produce very significant perceptual relations, but as we proceed, much larger time-stretching will present less and less surprising perceptual information. **(Sound example 11.24).**

If we wish to preserve the onset characteristics of a grain time-frame sound, we must use time-variable spectral time-stretching (*spectral time-warping*) (see above) or simply, perhaps, use the original grain as a superimposed onset for time-stretched versions of itself. **(Sound example 11.25).**

Conversely, a sound with continuation will usually have distinctive onset characteristics of grain dimensions. When we time-stretch such a sound we may choose to preserve the time-frame of the onset, hence preserving a key perceptual feature of the source, or we may time-stretch the onset too, producing, with large time-stretches, a dramatic change in the percept. Once again, some kind of mediation (through sounds with less time-stretched onsets) may be necessary to establish the perceptual link between source and goal sound, though the continuation of the two sounds may provide sufficient perceptual linkage between them. **(Sound example 11.26).**

As we have seen, grain-streams (and sequences) may be time-extended by *granular time-stretching by grain duplication*, by *granular time-stretching by grain separation* or by *spectral time-stretching*, which latter qualitatively transforms the grain (sequence) elements. These each have quite different musical implications. In particular, *granular time-stretching by grain separation* (in which the grains themselves are preserved while the intervening time-gaps are manipulated) will give us a clear percept of rhythmic variation (ritardando, accelerando, randomisation etc) as the internal pulse of the grain-stream (sequence) changes. In contrast, *spectral time-stretching* (or *brassage/harmoniser* type time-stretching) applied to *continuous sounds* is more akin to looking at time itself through a magnifying glass – the sound gets longer without any sense of the rhythmic slowing down of a pulse. When applied to a grain-stream these will usually create *both* a sense of rhythmic change *and* this sense of temporal extension. **(Diagram 2). (Sound example 11.27).**

Granular time-stretching by grain separation of a grain-stream will lead, ultimately, to a "slow" sequence of individual staccato events which may be re-sequenced in terms of onset-time (rhythm) pitch, or various spectral properties, or any combination of these. *Spectral time-stretching* or *harmoniser* time-stretching of a grain-stream, if it stretches the grain constituents by a large amount, will reveal spectral qualities and evolving shapes (morphologies) within the grain (or sequence) elements having their own musical implications. **(Sound example 11.28).**

The continuation of a sound, itself *spectrally time stretched* over a number of seconds (especially in complexly evolving sounds) may provide enough musical information for us to treat the extended sound as a phrase in its own right. As we have heard, time-variable spectral time-stretching (*spectral* *time-warping*) allows us to produce many versions of such a phrase with different internal time proportions and perhaps, spectral emphases, just as we might produce variants in the Hpitch domain, of a melodic phrase. Also, *time-warping* of a continuous sound can create a kind of forced continuation (see Chapter 5) in the spectral domain if applied to a complexly spectrally evolving sound. Similarly, undulating continuation properties will become slow glides with quite different musical implications to their undulating sources and we can alter these implications through *time-warping*. (**Sound example 11.29**).

The time-warping of rhythmically pulsed events provides us with an entirely new area for musical exploration. It is already possible to work with multiple fixed time-pulses (e.g. mutually synchronised click-tracks in different tempi, as in *Vox 3*) or with time-pulses mutually varying in a linear manner (as in the "phasing" pieces of Steve Reich, which generate new melodic, rhythmic or spectral percepts through finely controlled *delay*) but we might also produce mutual interaction between rhythmic streams which are themselves accelerating of decelerating, from time to time forcing the streams to pulse-synchronise in the same (possibly changing) tempo and in the same (or a displaced) pulse-grouping unit (bars coincide or not). (**Sound example 11.30**).

This device is used in *Vox 5* where three copies of an ululating voice are slightly time-varied in different ways and made to move differently in space. They begin in synchronisation, in a single spatial location, and reconverge to a similar state at the section end by appropriate inversions of the speed variations (see **Diagram 3**). Hence speed divergence and resynchronisation become elements in the emergence and merging of aural streams; or "counterstreaming", the time-fluid cousin of counterpoint. Note, however, that the sound constituents of two time-varying counter streams do not have to be the same, or even similar. (**Sound example 11.31**).

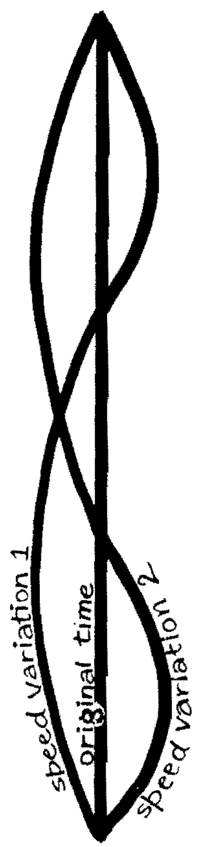
In fact, if the stream constituents are identical, are synchronised at some point and have slightly different time-stretch parameters, their interaction produces portamenti rising to or falling from the point of synchronisation (see Chapter 2 on Pitch Creation). (**Sound example 11.32**). In the previous case (**Sound example 11.31**) these precise-delay pitch-artefacts were avoided by carefully fading two of the three streams just before such artefacts would appear.

The interaction of time-varying pulse-streams bears the same relation to fixed-tempo polyrhythm as controlling pitch-glide textures does to Hpitch field pitch organisation.

Time-stretching may lead to *spectral fission* (see Chapter 3), e.g. Spectral time-stretching of noisy sounds by large amounts may produce sounds with gliding inharmonic spectra (**Sound example 11.11**) - or to *constructive distortion* e.g. *waveset time-stretching* produces pitches bead streams (**Sound Examples 11. 7 & 11.8**) and *spectral time-stretching* reveals window-stepping in *spectrally traced* sounds (**Sound Example 11.16**). In both cases a long time-stretched goal sound may not be perceptually relatable to the source and will require mediating sounds (e.g. less time-stretched versions of the source using the same technique) to establish a musical connection.

Throughout this section we have talked only of time stretching but the same arguments may be applied, in reverse, to time contraction. In particular, phrase structure may be compressed into grain-stream (a comprehensible spoken sentence can become a rapid-fire, spectrally irregular sequence of grains) and continuation compressed into the indivisible qualitative percept of grain. Once again, if a perceptual connection between source and goal is required (to build musical structure), mediating sound types may be necessary.

DIAGRAM 3



TIME-STRETCHING OF TEXTURE-STREAMS

We have left the discussion of texture-streams until the last because it introduces further multi-dimensionality into our discussion of time-stretching. We have already encountered a two dimensional situation with grain-streams. A grain-stream may be *granular time-stretched* by *grain separation*, or by *grain duplication*. The first process reduces the pulse-rate (or density) of the grain-stream, the latter does not. With texture-streams the situation is even more complex.

We may distinguish three distinct approaches to time-stretching a texture-stream. In the first we treat the texture-stream as an indivisible whole and time-stretch it. We may do this by *spectral time-stretching*, thereby stretching all the texture constituents, and hence very quickly producing a radical spectral transformation of the percept. All the perceptible time-varying field properties of the texture-stream will thereby be time-stretched e.g. loudness trajectory, pitch-band width change etc. The revelation of the inner structure of grains may even alter the field percepts (e.g. noise elements becoming inharmonic sounds, or Hpitches appearing as gliding pitches) of the source sound. (**Sound example 11.33**).

Waveset time-stretching will produce surprising and unpredictable artefacts when applied to texture streams as the zero crossing analysis will confuse the contribution of various distinct grains to the overall signal. With a stereo texture, waveset duplication may be applied to each channel independently, producing arbitrary phase shifts between the channels, as well as the aforementioned artefacts. (**Sound example 11.34**).

Brassage techniques with above-grainsize segments, using regular segment size and zero search-range (see **Appendix p44**) will quickly destroy the unpatterned quality of the texture-stream, as brassage repetitions introduce a "spurious" order into the goal sound. Brassage will work better at preserving the inherent qualities of the texture-stream if we use a large enough segment size to capture the disorder of the texture-stream and a large enough range to avoid obvious repetition of materials. However, too large a range will begin to destroy any time-varying order in the field characteristics of the stream (e.g. directed change of the Hpitch field, loudness trajectory etc). (**Sound example 11.35**).

Assuming we have fine control of the *texture generation* process we could, in fact, separate out some of these field properties e.g. time-stretching a dynamically flat version of the texture, then reimposing the original loudness trajectory in exactly the same time-frame as in the source texture-stream. We will discuss this parameter separation further, below. (**Sound example 11.36**).

The second approach to time-stretching a texture-stream would be *granular time-stretching by grain separation*, as with grain-streams and sequences. However, because of the mutual overlaying of grains (or larger constituents) in a texture-stream, there is usually no simple way we can achieve this. It can only be done, in general, by returning to the *texture generation* process and altering the event-onset density parameters. To achieve an integrated time-stretch of this sort, any time-varying field properties (Hpitch field change, loudness trajectory, formant change etc.) would need to be similarly time-stretched in the generating instructions. We could, however, choose *not* to alter these features of the stream. In this way, we may create a goal sound which appears less event-onset dense than the source sound but not perceptually time-stretched in any meaningful sense. (**Sound example 11.37**).

This suggests the third approach to time-stretching. In this case we may generate events at the original density, but for a longer time, then impose time-dilated field-variation parameters on the stream. Here the loudness trajectory, the pitch-range variation, the formant changes, the transitions to noisiness etc would move more slowly, but the event-onset density would remain as before. This is the textural equivalent of *granular time-stretching by grain duplication*. (**Sound example 11.38**).

We here begin to touch upon interesting music-philosophical ground. For in this last case, the texture-stream percept is clearly not a "unique" sound-event, in the sense which we spoke of this in Chapter 1. The texture-stream is an example of a class of sounds with certain definable time-varying properties, just as a *note*, in traditional music, is a representative of a class of sounds with certain definable stable properties. It is the musical context which focuses our attention upon particular properties, or groups of properties of a sound, or on its holistic characteristics. Composition focuses perception on what is being perceptually organised through time. Or rather it does this so long as it is aware of what *can be* perceived and what *will be* perceived in the resulting musical stream.

We have hence given three quite different definitions of time-stretching a texture-stream. If we include the possibility of *spectrally time-stretching* the texture-components prior to generating the texture-stream, we may imagine another option in which the texture constituents are *spectrally time-stretched* (this time-stretching itself changing from constituent to constituent, as we proceed timewise through the texture) while (as far as is possible) the temporal evolution of density and field characteristics remain unchanged. This might better be regarded as a time-varying spectral-type/duration-type transformation of the source texture.

In fact we can time-stretch event-onset-separation-density variation, event-duration variation, overall loudness trajectory, pitch-range variation, evolution of the spectral contour (formant evolution) etc. etc. all *independently of one another*. Time has thus become a multi-dimensional phenomena within the sound percept and we may choose amongst the many compositional options available to us.

CHAPTER 12

INTERPOLATION

WHAT IS INTERPOLATION ?

Any perceptually effective compositional process applied to a sound will produce a different sound. However, if the process is sufficiently radical (extreme *spectral stretching*, or *time warping* or *filtering* or randomisation of perceptual elements etc) we will produce a sound which we recognise as being of a different *type*. I will try to define this more clearly below.

For the moment, we also note that we can, by using a similar compositional process, create a whole set of sounds, whose properties are intermediate between those of the source and those of the goal. We will describe this mediation by progressive steps as *static interpolation*.

Alternatively, provided our sound is sufficiently long, we may gradually apply a compositional process changing the value of various parameters through time, e.g. we may gradually *spectrally stretch* the harmonic spectrum of an instrumental tone until we reach, through a continuous process, a complex and inharmonic sound. Or we may gradually add vibrato to a relatively short term, pitch-stable sound (e.g. dense traffic) so that it eventually involves such extreme and rapid swings of pitch that the original percept is swallowed up by the process. In these cases we have a process of *dynamic interpolation* taking place through the application of a continuously time-varying process.

Clearly, we can apply this kind of reasoning to all compositional intervention and all compositional processes might be described as so many sophisticated variants of interpolation. However, in this chapter we will deal largely with compositional processes which interpolate between two (or more) *pre-existing* sounds. And, in a similar way, we will discuss both static and dynamic interpolation between these sounds.

It is important to understand that in this case we wish to achieve some sense of perceptual *fusion* between the two original percepts - a mere superimposition of one over the other is not acceptable as an interpolation. This is discussed more fully below.

IS RECOGNITION IMPORTANT ?

A significant factor to define, when discussing the idea of interpolation, is the recognition of source and goal sounds in the process. What distinguishes interpolation from mere variation is our feeling that we have moved away from one type of sound and arrived at a different type. This percept is most readily understood when the source and goal sounds are recognisable in some referential sense. The source is a trumpet, the goal is a violin: the source is the sea, the goal is a voice: the source is spoken English, the goal is spoken French: or even the source is a singing voice, the goal is clearly not a voice. Interpolation may pass directly from one recognition (voice) to another (bees) or it may seek out an ambiguous ground in which two recognition concepts conflict or cooperate within the same experience (the talking sea).

However, a percept of interpolation is possible without such clear referential clues where there is a distinct change in the sense of physicality or causality of the source (see Chapter 1).

Hence, by greatly *spectrally time-stretching* a voice, or a flexed metal sheet sound, and then imposing a rapid series of hard-edged loudness trajectories on the resulting continuum, we move from the sense of forced continuation of an elastic medium, to a sense of striking a hard inflexible medium. Both physicality and causality have been altered. (**Sound example 12.1**).

Hence modifications to onset characteristics, the rate of spectral change (and elsewhere to the irregularity-regularity of sequencing etc) alter our intuitive type-classifications of the sounds we hear. When compositional processes move sounds across these boundaries, we create the percept of interpolation.

MEDIATION : AMBIGUITY : CHANGE

Before we go on to discuss compositional methods for achieving interpolation, it is worth considering why we might want to do it. There would seem to be at least three different motivations for musical interpolation and each motivation leads to a different emphasis in the way the technique is applied.

The first approach is aimed at achieving some kind of mediation in sound between distinct sound types. This approach may be heard in Stockhausen's *Gesang der Junglinge* where the 'pure' pitched singing voice of a young boy and pure (pitched) sine tones are mediated through a set of intermediate pitched sounds (sounds between 'boy' and 'sine tone').

This desire to mediate between the child's voice and a set of more 'abstract' (i.e. less source-recognisable) sounds, has a metaphysical underpinning (a religious conception of 'unity' in the cosmos), which pervades much of Stockhausen's musical thought. The mediation is not achieved through dynamic interpolation (technologically almost impossible at the time), nor through a clear progressive movement from one sound-type to the other, but in the sense that the piece is grounded in a field of sound types which span the range 'boy's voice' to 'sine tone'. These are sequentially articulated according to an entirely different logic (a serialist sequencing aesthetic), which also governs the rest of the musical organisation in the piece.

A second approach to sound interpolation stresses the ambiguous implications of the sounds thus created. Roger Reynold's has used interpolations between a voice speaking a Samuel Beckett text in English, the same voice speaking the text in French and the sound of brass instruments. Interpolation takes place in two dimensions, between English and French on the one hand and between voice and instrument on the other. The composer focuses on the cusp of the interpolations, where we are most undecided about whether what we hear is English or French, voice or instrument. This approach also has its own metaphysical implications, if of a more secular variety. Technically the aim (and difficulty) here, is to achieve a percept which is capable of these dual interpretations without entirely losing 'source credibility' (i.e. is it anything at all that we can recognise?). This can be particularly difficult, even with the most advanced technology.

The third approach focuses upon the process of change itself. In *Vox 5* the transformation voice->bees aims to achieve clear recognition of both source and goal, and a seamless transition from one to the other without any intervening artefacts which might suggest some other physicality/causality, or even betray the technical process involved. **(Sound example 12.2)**.

Moreover the dynamism of the change is a crucial parameter; in the voice->bees example, the voice almost melts slowly into the bee swarm; other transformations in *Vox 5* are quicker and more forceful, suggesting a generative energy in the vocal source, spitting or throwing out the goal sounds – and this dynamism is enhanced by spatialisation in the sense of spatial motion, or the emergence of stereo images from a mono source. Here, the technical problems are those of seamlessness in the spectral transition and achieving the right dynamism, especially as interpolation recognition often requires a relatively long time-frame in order to work smoothly. Musical context can play a vital role here.

INBETWEENING

The most obvious way to achieve some kind of interpolation between two sounds would seem to be to *mix* them in appropriate (relative loudness) proportions. However, as we know from our everyday experience this almost never creates perceptual fusion of the two sources. Our brain manages to unscramble the many sound impinging on our ears at any time and to sort them into separate sources. We hear mix, not fusion. This is due partly to the ear's sensitivity to onset synchronicity (or the lack of it) and partly to the parallelism of micro-fluctuations of the components in any one source, at the same time being different from that in other sources.

Successful interpolation by mixing can only be achieved if we can defeat either, or both, of these. In sounds with continuation, the very precise (to the sample) synchronisation of attack (*onset synchronisation*) can achieve an instantaneous fusion of the aural image which is however immediately contradicted by the continuation of the sounds in question. **(Sound example 12.3)**.

To achieve a completely convincing sound intermediate between two others, we must work with sounds whose continuations are (almost) identical, or with grain time-frame sounds (which have no continuation). In the latter case in particular, we can achieve good intermediate percepts simply by *mixing*, but not necessarily. The ideal case is one in which the source sound is transformed into the goal sound by a distortion process that retains the duration and general shape of the source down to the level of the wavecycle or waveset (see *destructive distortion* in Chapter 3). Superimpositions of these two sounds in various proportions will create convincing intermediate states (*inbetweening* : **Appendix p46**). **(Sound example 12.4)**.

Where the sounds have continuation, this process may be unsatisfactory. Consider, for example, the goal and source sounds of the sequence "Ko->u" to "Bell" from *Vox 5*. **(Sound example 12.5)**.

If we began with the goal and source of this sequence and merely mixed them in various proportions, we would not achieve a satisfactory set of intermediate sounds. The process of *spectral stretching*, successively applied, has gradually separated the spectral components (and hence altered the spectral structure) to a point where they will not simply fuse together again by *mixing*.

The difficulties here might be compared with those in the process of in-betweening in animation. Here the directing artist draws key frames for the film, and assistants (or computers) do the various in-between drawings such that, when all are combined frame-by-frame on the film, realistic movement will be created. In-betweening may not be as skilful an occupation as originating the key frames but it is by no means elementary. Simple point-to-point linear interpolation will, in general, not work.

INTERLEAVING AND TEXTURAL SUBSTITUTION

A different approach to this problem of static interpolation is to, in some sense, interleave the data from our two sounds. One way to do this is to interleave analysis windows obtained from the spectral analysis of two sounds.

When we produce a spectral analysis of a time-varying sound, we divide the sound into tiny time-slices, or windows, and analyse the spectrum in each window (using e.g. the *fast fourier transform*, see **Appendix pp2-3**, and Chapter 1) in order to follow the temporal evolution of the spectrum. This procedure is the basis of the *phase vocoder* (see **Appendix**).

Having produced such an analysis of our two different sounds we may interleave alternate windows from each sound, preserving the original time-frame, or simply interleave windows as they are, creating a goal sound as long as the sum of the original sounds' lengths (see Appendix). This procedure can result in a kind of closely welded 'mix', if not exactly a fusion of aural percept. We may also choose to interleave pairs (or larger groups) of windows. With a sufficiently large grouping of windows, this will, of course, produce a rapid regular oscillation between the two source sounds. **(Sound example 12.6)**.

A similar percept can, however, be more easily achieved through *brassage* where we also have control over the segment size and are not therefore tied to regular oscillations. Two or more sources can be brassaged together (*multi-source brassage*) with control of segment size, search range, pitch shift, spatialisation of segments etc, just as in single source brassage. This approach perhaps works best in producing a *process-focussed transformation* (see Chapter 1) (e.g. granulated or pitch-spread) in which the two (or more) sources 'peer through' the distorting grill of the process simultaneously. The process's artefact grill disguises the lack of true fusion of the sources – both are an integral part of the resulting 'mince'. **(Sound example 12.7)**.

We can go one step further and attempt to integrate the wave-cycles, or wavesets, of the two sounds. Interleaving wavecycles or wavesets will produce spectrally unpredictable sidebands; two interleaved sinewaves of different wavelengths will produce a sideband whose wavelength is the sum of the originals, but with complex signals the process is likely to result in radical mutual distortion. This becomes a mutual transformation process rather than a true interpolation (*waveset interleaving*). **(Sound example 12.8)**.

Similarly, we may impose the waveset (or wavecycle) shape of one sound on the waveset (or wavecycle) length of the other. With simple sources, e.g. square wave to sine wave, this will force the pitch of the latter on the spectrum of the former. With complex sources, using wavesets, the results are once again complex and perceptually unpredictable. Again, we have mutual transformation, but not true interpolation (*waveset transfer*). **(Sound example 12.9)**.

If we have a very rapid sequence or a texture-stream we may achieve a dynamic interpolation between two quite distinct states by careful element substitution, e.g. we might start with a texture-stream of vaguely pitched noise-granules spread over a wide pitchband and, through gradually tighter *band pass filtering* of the elements themselves, focus the pitch of the granules, while simultaneously narrowing the pitchband. In this way, we can force a broad band noise granule stream onto a single pitch which we might animate with formant-glides and vibrato articulations reminiscent of the human voice (*granular synthesis* : see Appendix). **(Sound Example 12.10)**.

VOCODING AND SPECTRAL MASKING

The distribution of partials in a spectrum (spectral form) and the spectral contour (formants) are separable phenomena and impinge differently on our perception. As discussed in Chapter 3 the spectral form creates our sense of harmonicity-inharmonicity, while the spectral contour contributes to our formant, or 'vowel' perception. If we therefore have one source with a clear articulation of the formants (e.g. speech), and another source which lacks significant formant variation (e.g. a flute, the sea), we can impose the formant variation of the first on the spectral form of the second to create a dual percept (*vocoding*). As the spectral contour defining the formants must have something on which to 'grip' on the second source, this process works best if the second source has a relatively flat spectral contour over the whole frequency spectrum. (see **Diagram 1**).

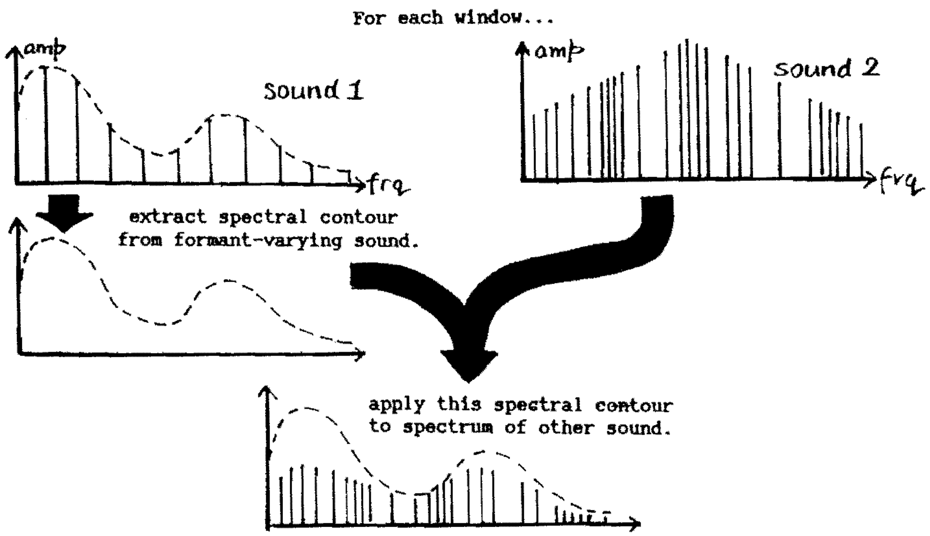
In speech analysis and synthesis, spectral contour data is recovered and stored as data defining a set of time-varying filters, using a process known as *linear predictive coding*. The speech can be reconstituted by driving a generating signal through these time-varying *filters*. A sequence of constant squarewave type buzzes of appropriate pitches (for voiced vowels and consonants) and stable white noise (for noise consonants and unvoiced speech) played through the time varying filters can be used to reconstitute the original speech. The process can also be used to reconstitute the speech at a different pitch (change the pitch of the buzz), or speed (change the rate of succession of the filters and the buzz/noise), or to change the voiced-unvoiced characteristics (choice of buzz or noise).

For interpolation applications, the signal we send through the filters will be our second source for interpolation (i.e. the one that's not the voice!). It will usually not have a flat, or even a stable, spectrum but we can enhance the formant transfer process by 'whitening' the second source, i.e. by adding noise discreetly to the spectrum in frequency ranges where there is little energy in the source. For obvious reasons, this process is most often used to interpolate between a voice and a second source and is often called *vocoding*. It should not be confused with the *phase vocoder*. (Similar vocoding procedures may, however, be attempted, using spectral data extracted by phase vocoder analysis).

This type of interpolation might also be applied progressively so that 'voiceness', or lack of it, emerges out of a continuing non-vocal sound. **(Sound example 12.11)**.

Making the sea 'talk' may be a sophisticated process of control of spectral contour evolution. However, the inverse process, making 'talk' like the sea, can be envisioned much more simply: a very dense texture of unvoiced speech to which an appropriate wave-breaking-shape loudness-trajectory is applied (using a choral conductor!). If the spectral type of the sounds within the texture stream is made to vary appropriately (e.g. lack of sibilants initially, 'f' and 'sh' sibilants at the wave peak, 's' sibilants with high formants for the undertow) we can create a dual percept with no electronic technology whatsoever. We might then proceed to vocode a recording of our 'sea' construct – voices within voices.

DIAGRAM 1



Another way in which we can achieve interaction between the spectra of two or more sounds is via ***spectral masking***. Here we construct a goal sound from the spectra of two (or more) source sounds by selecting the loudest partial on a frequency-band by frequency-band basis for each time window. (See **Appendix**). If one of our sources has prominent high frequency partials, it may mask out the high frequency data in the other source(s) and the high frequency characteristics of the masked sources will be suddenly revealed if the first source pauses, or gets quieter. Hence aspects of the spectra of two (or more) sources may be played off against each other in a contrapuntal interaction of the spectral data. This technique should, in general, be regarded as a form of spectrally interactive *mixing*, rather than interpolation in the true sense. However, if our two (or more) source sounds have stably pitched spectra which are strongly pitch-related, alterations in the loudness balance of source sounds will produce interpolated spectral states. With voices, or other sources with variable formants, this interaction will be particularly potent. (**Sound example 12.12**).

SPECTRAL INTERPOLATION

The most satisfactory form of dynamic interpolation is achieved by interpolating progressively between the time-changing spectra of two sources. (*spectral interpolation*). This process is used extensively in *Vox 5*.

In **Appendix p32** each sound is represented by a series of frequency domain analysis windows. The information in these windows changes, window by window, for each sound. Due to the nature of the analysis procedure, however, the windows are in step-time synchronisation between the two sounds, where the time- step corresponds to the window duration.

We can now apply a process of moving the amplitude (loudness) and frequency values in window N of sound 1 towards the values in window N of sound 2. If we do this progressively so that in window N+1 we can move a little further away from the values in sound 1 in window N+1 and a little closer to those values in sound 2 in window N+1, then in windows N+2 etc, the resulting window values will move progressively from being close to those in sound 1 to being close to those in sound 2 and the resulting sound will be heard to move gradually from the first percept (sound 1) to the second (sound 2).

It is important to understand that we are interpolating over the *difference* between the values in successive windows. We are moving gradually away from the *current* value in sound 1 towards the *current* value in sound 2, and not from the original values (back in window N) of sound 1 towards the ultimate values (onward in window N+K) of sound 2. The latter process, being a linear translation between two static spectral states, would produce merely a spectral glide perceptually disconnected from both source sounds. (**Sound example 12.13**).

Our process, in contrast, is moving from the 'wobbling' of one spectrum into the 'wobbling' of the other. For this very reason, the interpolation tends to be perceptually smooth. *Mixing* sounds normally fails to fuse them as a single percept because the micro-fluctuations within each sound are mutually synchronised and out of sync with those in the other sound. For this reason, a cross-fade does not produce an interpolation. In our process, we are effectively interpolating the micro-fluctuations themselves. (listen to **Sound example 12.3**).

There are a number of refinements to the interpolation process. We may interpolate amplitude data and frequency data separately (at different times and/or over different timescales) and we may interpolate in a linear or non-linear fashion, and we must choose these parameters in a way which is appropriate to the particular pair of sounds with which we are working. (see **Appendix p33**).

There are also perceptual problems in creating truly seamless interpolation between two recognisable sources. First of all, source recognition takes time and we need enough time for both source and goal to be recognised, as well as for the interpolation itself to take place. It is also important for the two sources to be perceptually similar (e.g. same pitch, similarly noisy), if the seamless transition is to be achieved without intervening artefacts which either suggest a third and unrelated physicality/causality in the sound, or even reveal the mechanics of the process of composition itself.

A more difficult problem is created by categorie switching in perception, e.g. in the transformation voice->bees, we tend to perceive - "a voice - a voice? - a voice??? - no, it's bees!" - there is a sudden switch as we recognise the goal percept. It may be necessary, to achieve a truly seamless transition, to create perceptual 'false trails' which distract the ear's attention sufficiently at the point of maximum uncertainty for the transition to be accepted. In the voice->bees transition in *Vox 5*, a very high-register, low-level part of the spectrum undergoes a slide at the 'moment of maximal doubt' in a way which is not consciously registered but seems sufficient to undermine the sudden categoric switching which had not been overcome by other means.

This type of spectral process can also be used for static interpolation, creating a set of sounds spectrally intermediate between source and goal. If we also allow progressive *time-stretching* (so that duration of source sound and goal sound can be matched through interpolation) and we place no aesthetic restrictions (such as the goal of 'seamlessness' in the examples above) on the intervening sound-types, it should be possible to produce a set of spectral intermediates between any two sounds - with one word of caution! Equal small changes in spectral parameters, need not lead to equal small perceptual changes. In fact a slight deviation in spectral form (e.g. harmonicity) may have a dramatic perceptual result.

Achieving a successful interpolation is about creating convincing, small *perceptual* changes in the resulting sounds - not about the internal mathematical logic that produces them. To achieve an approximately linear sense of interpolation along a set of sounds may require a highly non-linear sequence of processing parameter values. The proof of the mathematics is in the listening.

SPATIAL CONSIDERATIONS

Spatial perception may also be an important factor in creating convincing static or dynamic interpolation. In the crudest sense, imposed reverberation can help to fuse a sound-percept by giving an impression of 'spatial integrity' (this sound was apparently produced at a single place in a given space). More profoundly, spatial streaming will tend to separate a fused image; in a sequence or texture-stream, if one set of elements moves left and the other right, we experience aural stream dissociation.

A movement from mono to true stereo can, however, be used to enhance, or 'dramatise' the process of dynamic *spectral interpolation*. In *Vox 5* most of the voice->other interpolations start with a voice in mono at front centre stage and interpolate to a stereo image (crowd, bees etc) which itself often moves off over the listeners heads. Spatialisation and spatial motion hence reinforce the dynamism of the transition.

CHAPTER 13

NEW APPROACHES TO FORM

BEYOND SOUND-OBJECTS

There is still the danger of regarding sound–composition as a means to provide self–contained objects which are thenceforward to be controlled by an external architecture of Hpitch and rhythm along traditional lines. In particular, MIDI technology (1994) makes this option so easy and other approaches so roundabout that it is easy to give up at this stage and revert to purely traditional concepts for building large–scale form.

Furthermore, I do not wish to decry traditional approaches to musical form–building. On the contrary a broad knowledge of ideas about melodic construction and evolution, rhythmic organisation, HArmonic control, large–scale musical forms in general, etc, etc, from many different cultures (both "serious" and "popular") and historical periods, should underpin compositional choice. However, a compositional practice confined to this, in the late Twentieth Century, will inevitably be limited. I would particularly stress a detailed appreciation of singing styles and styles of declamation from around the world's cultures as a way to appreciate subtleties of sonic architecture and chemistry which are often missing or excluded from Western art music practice. Similarly, a study of the World's languages reveals the great range of sound materials that enter into everyday human sonic articulation somewhere on the planet.

The aim of this chapter is to suggest extensions to traditional ways of thinking. Extensions which are grounded in sound–composition itself. These may be used to complement or to replace traditional approaches depending on the sound context and the aesthetic aims of the composer.

MULTIDIMENSIONALITY

Given the priorities of notated Western Art Music, and the fact that these priorities are constructed into the instrument technology (from the tempered keyboard, or keyed flute, to the MIDI protocol) it is easy to view music as a two dimensional (Hpitch/duration) structure, "coloured–in" by sound. The very two–dimensionality of the musical page reinforces the notion that only two significant parameters can be precisely controlled – using horizontal staves and vertical bars. Sound composition involves rejecting this simplistic hierarchisation.

Sounds may be organised into multi–dimensional relation sets in terms of their Hpitch and Hpitch–field, or their pitch and pitch–band, their vowel set, harmonicity type, onset density, vibrato–acceleration–type etc. etc. with each of these treated as distinct perceptible orderable parameters. Not all listable parameters are perceptually separable in every case, e.g. vibrato acceleration type may be imperceptible when onset–density is very high. But in all definable sonic situations we have many sound parameters at our disposal.

We may also organise sounds in terms of their overall holistic qualities. This approach is appropriate both for grain time–frame events where separable classes of properties may not be distinguishable and for sound sets created by progressive interpolation between quite different sounds (e.g. "ko–>u" to "Bell": listen to **Sound example 12.5**). This latter sound interpolation illustrates the fact that we can create scales of perception between definable points without parameterising our experience beforehand. From a mathematical viewpoint, in a multi-dimensional space, we do not need to create scales along any preferred axes of the space; we can create a stepped line in any direction.

We may then explore this multi-dimensional space establishing relationships between sounds (holistically or through different property sets) which are both perceptible and musically potent in some sense.

Given this multi-dimensional space, we can generalise the notion of "modulation" (in the tonal sense). We are already aware that reorchestrating Hpitched music can fundamentally change its affective character. Music transferred from vibraphones to shawms, for example, or within instrumental practice, a change in "expression" in the production of the sound stream (e.g. bowing attack, and/or vibrato continuation on violins).

The distinction made between "structure" and "expression" is in fact an arbitrary, ideological divide. All the changes to the sounds can be traced to structural properties of the sounds and the control of those structural properties. The distinction structure/expression arises from the arbitrary divide created by the limitation of notation. Features of sound we have, in the past, been able to notate with some exactitude (Hpitch, duration) are opened up to "rational" control (or at least rationalised explication) by composers and commentators. Those which remain invisible or vague in the notation are out of reach of this rationalisable control. They do, of course, remain under the intuitive control of the performer (see Chapter 1) but this type of control comes to have a lower ontological status in the semantics of music philosophy.

Sound recording and computer control destroy the basis for this dualistic view. The multi-dimensional complexity of the sound world is opened up to compositional control. In this new space of possibilities, reason (or rationalisation), must come to terms with intuition. With precise sound-compositional control of the multi-dimensional space, we can move from what were (or appeared to be) all-or-nothing shifts in sound-type to a subtly articulated and possibly progressively time-varying "playing" of the sound space. Moreover, we do not have to treat each parameter as a separate entity. We may group properties into related sets, or link the way one property varies with the variation of another (e.g. vibrato speed with depth) - and we may vary these linkages.

We are already familiar with such subtle articulations of a multi-dimensional sound space within our everyday experience. Consider the many affective ways to deliver a text, even where we specify no significant change in tempo or rhythm. The range of human intent, physiological, health or age characteristics, pre-existing physical or emotional condition (breathlessness, hysteria etc.), textual meaning (irony, accusation, information, questioning etc) which can be conveyed by multi-dimensional articulation of the sound space, is something we take for granted in everyday social interactions and in theatre contexts.

With precise sound-compositional control of this multi-dimensional situation for any desired sound, we can see that there is a significant and subtly articulable space awaiting musical exploration.

THE STRUCTURAL FUNCTION OF INTERPOLATION

It is instructive to examine the form of some sound composition phrases to illustrate this multi-dimensional approach. The sound examples here are taken from *Tongues of Fire* (1992-94). In **Sound Example 13.1** we begin with a vocal sound whose tail is *spectrally time-stretched* with a gentle *tremolo* at its end. The prior context is that of voice sounds, Voice sounds themselves are only recognisable as such through a complex interaction of properties (pitch-tessitura, pitch-glide speed and range, formant and formant-glide set, noise types, general rate and semi-regularity of sequencing etc.).

After the first downward portamento we arrive at a section based on variants of this time-stretched voice-tail. Here the sliding inharmonic sound falls or rises or is *vibrato* or *tremolo* articulated in numerous ways. Each event in this segment begins vocally, but the spectrally time-stretched tail extensions are linked through their (time-evolving) spectral type; pitch and loudness are the principle articulating parameters.

This leads us to a varied recapitulation of the falling portamento with tremolando idea, but now the loudness trajectory of the tremolando cuts so deep that the originally continuous sound breaks into a succession of wood-like onsets. This is a sonic modulation from voice to 'wood' and is akin to a key change in the tonal system. However, this is only a passing modulation. We switch back again into a section of variants on the time-stretched tail of the vocal sound.

This section ends with a true sonic modulation, the ritardandoing wood events firmly establishing a new sonic area, no longer vocal, but 'wood-like'. It is interrupted by a surprising shift back towards the voice-tail, but, in fact this octave stacked, onset-enveloped version of the tail occurs without the vocal initiator, thus underlining the sonic "key change" which has taken place. As this event clears, the wood-like events rise in pitch and density, establishing a high granular texture. The voice has now been left behind. We are in a new sonic "key".

This sense of elsewhere is reinforced by a further pair of passing sonic modulations, as an accelerating wood-event (twice) transforms into a drum-like attack, itself colored by a very high frequency version of the granular texture, adding a cymbal-like presence to these onsets.

The section ends with yet another surprising sound-modulation as the vocal onset returns but (out of a set of other transforms) goes over into a sound like random vocal grit.

In my description of this event I have tried to emphasise the structural importance of sonority by comparing it to tonal structure. I am also stressing the structural function of interpolation as a continuum-domain analogue of tonal modulation. There are clearly important differences. The set of keys in the tonal system form a finite, discrete and cyclic set of possibilities (see *On Sonic Art*) and the tonal system is rooted in this underlying structure. The Sonic Continuum is neither discrete nor cyclic - but, by contrast, it is wonderfully multi-dimensional, and we still retain some sense of the 'distance' of one sound from another even if we cannot measure this in the way we can measure the distance around the cycle of fifths (see also the discussion of measured, comparative and textural perception in Chapter 9).

In this context, interpolation provides the sense of logical passage from one sonic area to another (as opposed to mere juxtaposition) which we also feel during tonal modulation.

We can also pick out other features of this sequence which help bind it together: the falling pitch shapes (and their rising inversion), the decelerating (and related accelerating) rhythmic patterns. What is significant is that many different dimensions of sound organisation can provide structural reference points. We no longer have the traditional hierarchy: Hpitch, duration, others.

In **Sound example 13.2** we begin with a sequence whose structure is rhythmically grounded (we hear varied repetitions of a rhythmic cell whose original material was vocal in origin: this fact is more apparent in the context of the whole piece than in the context of the example here) into which a resonant pitched event is interjected. As the sequence proceeds the units are given spectral pitch through successive *filterbank* filtering with increasing Q, and the structural focus partly passes over to the pitch domain. Once the 'fireworks' enter, the rhythmic patterning is lost in a dense texture, but the pitch-focus remains in the HArmonic field of this texture. In the foreground the 'fireworks' are linked by a new device, the falling portamento shape whose origin we hear when the *spectrally time-stretched* voice peers through the texture.

In **Sound example 13.3** we begin with a vocal texture increasing in density, which begins to rise in tessitura. As it does so the texture *white's out* to produce a noise band which is subsequently *filterbank* filtered to produce a multi-pitched inharmonic sound. The rising tessitura of the texture is first perceived as a secondary property (an articulation). But it soon becomes the binding structural feature of the ensuing section, as both noise bands and inharmonic sounds glide up and down.

It is the very multi-dimensionality of the sonic continuum which in some sense forces us to call upon different perceptual foci as binding elements at different times. I am not suggesting that one consciously composes in this way but that an intuitive sense of formal cohesion over an unhierarchised set of sonic properties may be important in sound composition.

FROM THE RATIONAL TO THE REAL

As I have argued elsewhere (*On Sonic Art*), the two dimensional grid of staves and barlines of the musical score reinforces a culturally received conception of the musical universe as lying on a grid, or lattice, of discreet values, particularly of Hpitch and of duration (but also instrument type). Notation hence deals with the "rational" in the mathematical sense of the term, i.e. that which can be counted, or that which can be expressed in terms of ratio: finite sets of Hpitches and the associated intervals, event-durations as multiples or divisions of a regular unit (e.g. a crotchet). The grid hides from us the reality of the musical continuum from which it is carved. But just as the integers and the rational numbers (fractions) are only special cases lying along the underlying continuum of the real numbers, so Hpitch sets and countable rhythmic units are special cases of an underlying continuum of frequency and duration values matched by the continua of formant types, spectral forms etc. etc.

Rationality, however, is in the ear of the beholder. For musicians accustomed to working in the Western tempered scale, there is nothing more simply rational than the intervallic division of the octave in terms of the logarithm of frequency by which we have come to measure pitch-interval. All semi-tones are equal and the fifth is simply 7 steps up a 12 step scale. Musicians from other cultures,

and devotees of various pre-tempered approaches to tuning (e.g. just intonation), however would beg to differ from this analysis.

For them the interval we know as the 5th derives from the second two partials in a harmonic spectrum. The ratio of their *frequencies* is exactly 3:2 or 1.5 . However, in the tempered scale, where larger intervals must always be expressible as exact multiples of the semitone (whose frequency ratio is 2 raised to the power of 1/12), the interval of a 5th is 2 raised to the power of 7/12, or *approximately* 1.498307077. In fact, in terms of frequency, the tempered scale 5th cannot be expressed as an exact ratio of any two whole numbers; it is not a rational number (in the mathematical sense). (It is certainly out of tune with the 'pure' fifth).

The tritone in the tempered scale, roughly corresponding to the mediaeval "devil in music", is even more instructive in this regard. For the tempered scale tritone corresponds to a precise frequency ratio of the-square-root-of-2:1. And the square root of 2 is perhaps the most famous non-rational number of all.

Pythagoras' discovery of the link between musical pitch relationships and the length ratios of vibrating strings spawned the pythagorean cult which linked mathematics, mystical numerology and music. However, it was the Greek discovery of the non-rationality of the square root of 2 which destroyed Greek arithmetic and led to an almost complete reliance on geometrical methods. For the square root of 2, the length of the diagonal of a square of unit side, is perfectly comprehensible geometrically. However as a *number* it cannot be expressed as the ratio of two other numbers. This was proved by the Greeks and seemed hopelessly baffling. The square root of 2 is simply not a rational number (a number expressible as a fraction).

It was only 2,000 years later, in the Seventeenth Century, with Descartes' development of coordinate geometry (linking geometrical and numerical thinking) and Newton's use of fluxions, (the origin of the calculus) to study accelerated motion, that mathematicians began to deal in a sophisticated way with the numerical continuum. And it was not until the Nineteenth Century that a rigorous definition of these new numbers, the very stuff of the continuum, was developed.

The numbers which form the continuum (and which include the rational numbers) are known as the Real Numbers. So, in the mathematical world, the rational appears as only a small part (in fact an infinitely small part) of the Real!

In a sense, Western musical practice has remained overawed by the Pythagorean fixation with the mathematically rational. The obsession with geometrical proportions in written scores may be seen as an aspect of the persistence of this tradition.

But the computer as a recording and transformational tool has altered our relationship to the musical continuum. Sampling is a means whereby the continuum of our sonic experience can be captured and subjected to rational calculation. The key to this is the existence of perceptual time-frames. Below a certain duration limit (as discussed in Chapter 1) we can no longer distinguish individual sound events, hence our perception of the continuity of experience can be recreated by piecing together extremely short, but stable, sound elements. Just as the 24 or 25 static frames of film or video create for us the illusion of seamless motion, the reconstruction of a sound from a windowed analysis, or the playing of samples through a digital to analogue converter, recreates for us a continuous sonic domain from what are fixed and countable values. The computer thus provides a link between the world of rational mathematics and the continuum of our sonic experience (so-called "floating point" arithmetic is a rational approximation of operations on real, in the mathematical sense, numbers).

It must also be said that, in dealing with the continuum mathematically on the computer, we both learn the necessity of numerical approximation and come up against the limits of human perceptual acuity which itself sets limits to the accuracy *required* for our calculations. The world of "pure" ratios in the sense of tuning systems, or temporal proportions, is seen for the idealisation that it is. We come face-to-face with the Real, the reality of human experience, and the realities of rational calculation.

Nevertheless, what before lay only within the intuitive control of physical action (which varies spatially and temporally over the continuum) in performance practice, is now recordable, repeatable, analysable, recalculable in a way not previously available to us.

SEQUENTIAL AND MORPHIC FORM

An important question is, then, to what extent can we distinguish and appreciate articulations of the continuum. Appreciation of musical performance, and subtle comprehension of the "innuendo" of spoken language, suggest we have a refined, if not well described, ability in this sphere.

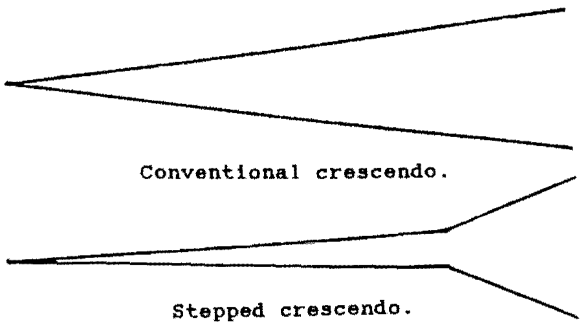
We will describe shapes articulated in the continuum as *morphic forms* in contrast with forms created by juxtaposing fixed values, which are *sequential forms*. Almost all Western Art Music *as it appears in the notation* is concerned with articulating sequential forms.

The invention of the orchestral crescendo by the Mannheim School of Symphonists seemed a major advance at the time. The crescendo is an example of the most elementary morphic form, a linear motion from one state (in this case, of loudness) to another. In a world of stable loudness fields (not taking into account the subtleties of loudness articulation in the performance practice) this was a startling development. But traditional notation gives no means to add detail to this simple state-interpolation. (More recent notational developments include stepped crescendi: see **Diagram 1**.)

Similarly, tempo variation, as expressed through traditional notation, cannot be given any subtlety of articulation. It is either happening at a "normal" rate (accel), slowly, (poco a poco accel) or rapidly (molto accel). Internal articulation of the rate of speed change is not describable (though some twentieth century composers, like Stockhausen, have attempted to extend notational practice to do this, in works for solo performer). Performance practice may involve the subtle or involved use of "rubato" (performance "feel") related tempo variation. It is interesting to note how "rubato" interpretation, associated with musical romanticism, is generally frowned upon in serious music circles. From a notation-focused perspective, it clearly does not adhere to the notated time information. But the fact is, we cannot notate this kind of subtle tempo fluctuation. Does this mean therefore it must be abandoned as an element in musical construction?

The subtle articulation of pitch, continuation and spectral formant properties can be observed in the vocal music practice of many non-European musical traditions, as well as within jazz and in popular musics. Musics handed down, at least partly, by aural tradition, can carry the morphic form information from one generation of performers to another. It does not get hived off into a separate domain of "performance practice" separated from stable-property notatables and subsequently devalued.

DIAGRAM 1



With computer control and the possibility of analytic understanding, we can imagine an extension of morphic control to larger and larger time-frames and to many levels and dimensions of musical experience. However, because of the existence of the text/performance separation in Western Art Music and the textual fixation of analysis and evaluation, it is difficult for music analytic thought to take on board morphic form. Such a vast literature of musical analysis exists, grounded in sequential properties. By contrast, there has been no means to notate, let alone describe and analyse, morphic form.

The combined powers of the computer to record sound and to perform numerical analysis of the data to any required degree of precision immediately removes any technical difficulty here. For those willing to deal with precise numerical representations of sonic reality, a whole new field of study opens up. To those fixated by musical texts, however, this area will remain a closed book.

From a perceptual perspective, even in the most elementary case, a motion from state A to state B, we are able to distinguish articulations of the motion. For example, a pitch-portamento which rises steeply before slowly settling on the goal pitch, can be differentiated from a linear portamento and from one which leaves its initial pitch slowly before accelerating towards the goal. These different motion types can be observed to have different perceptual qualities even within very short time-frames. **(Sound example 13.4).**

Once we begin to combine morphic form over a number of parameters, we may observe the emergence of elementary morphic categories. Simple examples would be....

1. The descrescendoing rising portamento which accelerates away from a Hpitch, a "leading-from" or dispersal morpheme.

2. A rising portamento which slows as it approaches the final Hpitch and from which gradually emerges a widening vibrato (also making the percept louder) - a common gesture in Western popular music - a "leading to", or morphic anacrusis.

3. A sense of morphic stability achieved on a sustained tone (with jitter) with or without a fairly stable vibrato depth and rate, as compared to,

4. Morphic instability occasioned if the vibrato on the same note begins to vary arbitrarily in speed and depth over a noticeable range.

(Sound example 13.5)

Such formal groupings can be transferred to other parameters (loudness variation, change of cyclic loudness variation (tremolo rate), formant motion, harmonicity shift, density fluctuation etc. etc.). In other cases, simultaneous morphic development of parameters may lead in contradictory directions. A tone whose vibrato-rate gradually slows but whose vibrato depth simultaneously widens ceases to be perceived as a tone event and becomes a pitch-glide structure. We have a simple example of morphic modulation, taking us from (perhaps) Hpitch field percepts to pitch-glide field percepts.

As morphic forming is developed over larger time-domains, new musical formal structures may develop. In particular, different musical streams, undergoing internal morphic change, may interact with one another. We have the possibility of counterstreaming, the morphic flow equivalent of counterpoint

in the sequential domain. Stream interaction in an elementary sense can be heard in Steve Reich's phasing pieces and in the ululation section of *Vox-5* (see Chapter 11) but in neither of these cases is the material in the separate streams undergoing real morphic change (in *Vox-5* the streams move in space relative to one another).

Imagine, for example, a detailed control of several pitch-portamento lines over long time-frames, or a texture-stream which undergoes continual morphic development of density, granularity, spectral energy focus, pitch-band width, pitch-band-location, internal pitch-stream motion etc, and which can diverge into perceptually separate streams, perhaps moving separately in space, before uniting sonically and spatially at a later time.

Imagine also a theatre situation in which we see the quite separate evolution of two characters with entirely different, but internally consistent, goals and desires. Then external circumstances conspire to cause the two to meet briefly in a small room. The theatrical outcome is dramatic and we, as audience, have an inkling of how this collision of personalities might turn out, whereas the characters are unprepared for this chance encounter. After their meeting, both their lives are changed dramatically. From a broader perspective, this is also stream-counterpoint, the collision and mutual interaction of developing streams.

The possibilities in this domain remain almost entirely unexplored, a rich seam of development for any sound composer who can grasp the concepts and mathematical controls of morphic flow processes.

POSTLUDE : TIME & SPACE

The forms of living organisms are the result of continuous processes of flow (growth) - they are morphic forms captured in space. Even the calcified remains of sea creatures (e.g. the shell of the Nautilus) tell us about the processes of continuous growth which formed them, and any spatial proportions we may observe in the final forms are merely the traces of these continuous processes, time-flows revealed in space.

What strikes us as formally coherent about a tree is not some exact and static proportioning of branching points and branch angles, but the way in which a flexible (and environmentally sensitive) expression of some branching principle is realised through a process of branching flow. The overall state of a particular flow pattern (its result being the current state of the tree) links our perception of one member of a tree species with the other members of the same species.

Unlike the growing tree, a musical event never fixes its time-flow evolution in a spatial object, it is intrinsically evanescent and a musical score is not the musical experience itself, but a set of instructions to generate this aetherial substance, just as the genetic code is a set of instructions to grow an organism. The symmetries of the DNA molecule are not, in any sense, copies or analogues of the symmetries of the organism. Similarly, we cannot assume that spatial order in a musical score corresponds to some similar sense of order in our musical experience of time-flow.

The way in which we divide experience into the spatially codifiable and the flowing governs much more than our attitude to music. Human social life itself is an experience of interactive flow, or growth, amongst human individuals. Some people find all flow (growth, change) unsettling and seek to codify experience, professional practice, social order and even human relationships in strict categories. Others see categoric divisions as temporary pens in which we attempt to corral the flow of experience and history.

Similarly, some artists seek out exact rational proportions, golden sections or, more recently, fractality, in everything. Perhaps it may seem to reveal the hand of the ultimate designer in all things. or, as with Plato, reality may be conceived as a hidden world of fixed Ideals to which everyday reality corresponds with varying degrees of success. In such a world, the score (the text), or the compositional idea or method, may be viewed as a musical Ideal and its realisation in performance an adequate or poor reflection of this Ideal object. Reality is set in stone, or print, in the exact proportions of architecture, the unchanging fixity of the text, the codification of the artist's method.

For artists like myself, reality is flow, growth, change and ultimate uncertainty. Texts, from the religious to the scientific, are our mortal attempts to hold this flow in our grasps so we can understand it and perhaps control it a little. Compositional methods are, like engineering methods, subject to test, and ultimately to possible failure.

For a certain time a procedure, an approach, a theory, works, until we discover new facts, new symmetries, new symmetry breaking, and those texts have to be revised or rewritten, the proportions of our architecture reconsidered, the appropriateness of our methods reassessed in the light of experience.

For me, sound–composition, in seeking the measure of the evanescent flow of sonic events, attempts to grapple with the very essence of human experience.

APPENDIX 1

(Page numbers refer to the diagrammatic Appendix 2).

***** A *****

ACCELERANDO

An increase in the speed at which musical events succeed one another.

ACCUMULATION

Sound which 'gathers momentum' (increases in loudness and spectral content) as it proceeds.

ALL-PASS FILTER 9

AMPLITUDE

A scientific measure of the strength of a sound-signal. This normally reveals itself in perception as the loudness of the sound, but amplitude and loudness and not equivalent. In particular, the sensitivity of the human ear varies from one frequency range to another, hence Amplitude and perceived loudness are not identical. However, in this book, the term Loudness is used, instead of Amplitude, in the text, wherever this does not cause any confusion. Diagrams however usually refer to Amplitude.

ANALYSIS

In this book analysis almost always refers to Fourier Analysis, the process of converting the waveform (time-domain representation) of a sound into its spectrum (frequency-domain representation). Analysis of musical scores is the process of determining the essential structural features of a composition from its notated representation in a score. This notion might be generalised to include any process attempting to uncover the underlying structure of a piece of music.

ARTICULATION

The (conscious) changing/moving of a property of a sound, usually in a way bound by conventions (of language, musical practice etc) but often not appearing as such in any associated notation system.

ATTACK

The onset portion of a sound where it achieves its initial maximum loudness.

***** B *****

BAND PASS FILTER 7

BAND REJECT FILTER 7

BAR

A grouping of musical duration–units. The bar–length is measured in some basic musical unit (e.g. crotchets, quavers) for which a speed (tempo) is given (e.g. 120 crotchets per minutes). In most musics, bar length is regular for long stretches of time, and bar length is one determinant of perceived rhythm.

BRASSAGE 44–45

A procedure which chops a sound into a number of segments and then resplices these together tail to head. In the simplest case sounds are selected in order, from the source, and respliced back together in order. However there are many possible variations on the procedure. Brassage may be used for changing the duration of a sound, for evolving montage based on a sound–source, and for many other musical applications. See also GRANULAR RECONSTRUCTION.

BREAKPOINT TABLE

A table which associates the value of some time–varying quantity (e.g. pitch, loudness, spectral stretch etc.) and the time at which that value is reached. e.g....

(time)	(value)
0.0	2.7
1.3	2.2
1.7	2.1

This table allows us to describe the time–varying trajectory of the quantity, and a musical program may read the data in the table, interpolating between the given values where necessary, to determine what to do at a particular time in the source–sound.

***** C *****

CADENCE

A recognised device in a musical language which signals the end of a musical phrase, section or piece.

CAUSALITY

The way in which the sound was apparently initiated. NOT the actual cause of the sound. Was the source apparently struck, rubbed, shaken, spun etc. ??

CHANGE RINGING

A form of bell–ringing in which the order in which the bells are sounded is permuted in specific ways.

CHANNEL

Channel is most often used in this book to refer to an analysis channel. We derive the spectrum (frequency domain representation) of a sound from its waveform (time–domain representation) by a process of analysis. In doing the analysis we must decide how accurate we would like to be. We may search for a partial in each block of 100 cycles per second (i.e. between 50 and 150, 150 and 250, 250 and 350 etc) or, more discriminatingly, in each block of 10 cycles per second (i.e. between 5 and 15, 15 and 25, 25 and 30 etc.). These search blocks are the channels of the analysis. Channels should not be confused with WINDOWS.

Channel is also used to refer to the right hand and left hand parts of a stereo sound (which can be viewed as two separate streams of digital information).

CHORD

A set of pitches initiated and sounding at the same time. Usually a set of pitches within a known reference set (e.g. the European tempered scale).

CHORUSING 45

A process of making a single musical source (e.g. a voice) sound like a group of similar sources all making the same sound (e.g. a chorus of singers singing the same pitch).

CHROMATIC SCALE

The european scale consisting of all the semitone divisions of the octave.

COMB–FILTER TRANSPOSITION........... 65

A quick method of achieving octave–upward transposition of a sound, where its pitch is known.

COMB–FILTERING 64

COMPRESSION 60

Reducing the loudness of a sound by greater amounts where the sound itself becomes louder.

CONSTRUCTED CONTINUATION

The extension of a sound by some compositional process (e.g. brassage, zigzagging).

CONSTRUCTIVE DISTORTION

A process which generates musically interesting artefacts from the intrinsic properties of the waveform or the (time-varying) spectrum of a sound.

CONTINUATION

Where sounds are longer than grains, we hear how the sound qualities evolve in time, (their morphology). These sounds have continuation.

CONTOUR

The shape of some property of a sound at one moment in time. In particular, the shape of the spectrum. This is often referred to as Spectral Envelope. However, Envelope is also used to describe the time-changing evolution of a property (especially Loudness). To avoid any confusion, this book reserves Trajectory for such time-changing properties, and Contour for the instantaneous shape of a property. The names of computer instruments may however use the term 'Envelope'.

The spectral contour describes the overall loudness contour of the spectrum at a single moment in time. But note that the spectral contour may itself evolve (change) through time.

CORRUGATION 60

CROTCHET = 120

An indication of the speed at which musical events succeed one another. Here the duration unit, crotchet, occurs 120 times every second. This speed is known as the Tempo.

CSOUND

A general purpose computer language which allows a composer to describe a sound-generating procedure (synthesis method) and ways to control it, in almost any degree of detail, and to define a sequence of events (a score) using the sounds generated, and which then generates the sound events thus defined. CSound is the most recent development of a series of such general purpose synthesis engines, and the one in most common use at the time of writing (Autumn, 1994).

CUTTING 40

***** D *****

DELAY 64

DENSITY

Describes the way in which a range is filled. Applies particularly to time ranges. A high-density texture has a great many events in a short time. Temporal density is a primary property of TEXTURE-STREAMS. The concept of density can also be applied to pitch-ranges.

DESTRUCTIVE DISTORTION

An irreversible transformation of the waveform of a sound, changing its spectral quality (the brightness, noisiness etc), rather than the pitch or duration. Distortion also implies the degradation of the sound (and irreversibility means that the original sound cannot be restored from the distorted version). Destructive distortion which preserves zero-crossing points can be musically useful. See WAVESETS.

DIPHONE SYNTHESIS

The reconstruction of speech (usually) by synthesising the transitions between significant phonemes (roughly speaking, vowels & consonants) rather than the phonemes themselves.

DRONE

A pitched or multipitched sustained sound which persists for a long time.

DUCKING 62

Means of ensuring the prominence of a lead 'voice' in a mix.

DURATION

The length of time a sound persists. Not to be confused with event-onset-separation-dration, which is the time between the start of successive sound events.

DYNAMIC INTERPOLATION

The process whereby a sound gradually changes into a different (kind of) sound during the course of a single sonic event.

***** E *****

ECHO 64

EDITING

The processes of cutting sounds into shorter segments or/and splicing together sounds or segments of sounds.

ENVELOPE

The loudness trajectory of a sound (the way the loudness varies through time) is often referred to as the envelope of the sound. Computer instruments which manipulate this loudness trajectory are usually called 'Envelope something'. Envelope is also used in the literature to refer to the time–changing variation of any property (we use the term trajectory) and even to the instantaneous shape of the spectrum (we use the term contour).

ENVELOPE CONTRACTION 60

ENVELOPE FOLLOWING 58

ENVELOPE INVERSION 60

ENVELOPE SMOOTHING 60

ENVELOPE SUBSTITUTION 59,61

ENVELOPE TRANSFORMATION 60

Musical transformations of the loudness trajectory of a sound.

ENVELOPING 59

EXPANDING 60

***** F *****

F5

The pitch 'F' in the European scale, in the 5th octave.

FAST FOURIER TRANSFORM

A very efficient computer algorithm for performing the Fourier Transform.

FF

Fortissimo = very loud.

FIBONACCI SERIES

Series of numbers in which each term is given by the sum of the previous two terms. The sequence begins with 1,1 and is hence..

1, 1, 2, 3, 5, 8, 13, 21, 34, 55 etc

The ratio between successive terms

1/1, 1/2, 2/3, 3/5, 5/8, 8/13, 24/21, 21/34, 34/55 etc.

approaches closer and closer to the Golden Section as one goes up the series.

FIELD

A set of values defined for some property. e.g. the pitches of the scale of G minor define a Field. Any pitch of the tempered scale lies either in that field or outside it. Any pitch within the pitch continuum lies close (in some perceptually defined sense) to that Field or not.

FILTER 7–8

An instrument which changes the loudness of the spectral components of a sound (even reducing some to zero).

FILTER BANK 7

FLUTTER–TONGUING

A way of articulating a granular sound on breath controlled instruments (woodwind or brass) by making a rolled 'r' in the mouth while producing a pitch on the instrument.

FORMANT 10

FORMANT PRESERVING SPECTRAL MANIPULATION....... 17

FOURIER ANALYSIS 2

The representation of the waveform of a sound as a set of simpler (sinusoidal) waveforms. The new representation is known as the spectrum of the sound.

FOURIER TRANSFORM

A mathematical procedure which allows us to represent any arbitrary waveform as a sum of elementary sinusoidal waveforms. It is used in Fourier Analysis. See also INVERSE FOURIER TRANSFORM.

FREQUENCY 3,4

A steady sound has a definite repeating shape, a waveform, and this waveform a definite length, which takes a certain time to pass the listener. The number of waveforms that pass in each second (the number of cycles per second) is known as the frequency of the wave. The frequency of the wave helps to determine the pitch we hear.

FREQUENCY DOMAIN REPRESENTATION 3

***** G *****

GATING 60

GOLDEN SECTION

If a straight line AB is cut at a point P such that...

AP/AB = PB/AP

this ratio is desribed as the Golden Section, & is approximately equal to 0.618

A - - - - - - P - - - - B

GRAIN

Sound having a duration which is long enough for spectral properties to be perceived but not long enough for time-fluctuation of properties to be perceived.

GRAIN-STREAM

A sound consisting of a rapid sequence of similar onsets.

GRAIN TIME-FRAME

The typical duration of a grain.

GRANULAR PROCESSES 56-57

Musical processes which preserve the grains within a grain- stream.

GRANULAR RECONSTRUCTION 73

A procedure which chops a sound into a number of segments and then redistributes these in a texture of definable density. The process differs from BRASSAGE in that the segments need not be rejoined tail to head. Granular Reconstruction of output density 1 is brassage.

GRANULAR REORDERING 57

GRANULAR REVERSAL 56

GRANULAR SYNTHESIS

A process, almost identical to granular reconstruction, in which very brief sound elements are generated with particular (time-varying) properties and a particular (time-varying) density, to create a new sound.

GRANULAR TIME SHRINKING BY GRAIN DELETION 56

GRANULAR TIME SHRINKING BY GRAIN SEPARATION 56

GRANULAR TIME STRETCHING BY GRAIN DUPLICATION 56

GRANULAR TIME STRETCHING BY GRAIN SEPARATION 56

GRANULAR TIME WARPING

Granular time-stretching or time-shrinking, which itself varies in time.

***** H *****

HARMONIC

In traditional European practice, harmonic means pertaining to harmony and relates particularly to tonal music. In sound composition, harmonic refers to the property of a spectrum where all the partials are multiples of some (audible) fundamental frequency. These partials are then known as harmonics, the spectrum is said to be harmonic and the sound has a single definite pitch. (See also INHARMONIC). These two usages are incompatible, so in the text we use HArmonic to refer to the traditional usage, and harmonic to refer to the sound-compositional usage.

HARMONIC FIELD

A reference frame of pitches. This might be thought of as a chord. All pitches in a texture controlled by a HArmonic Field will fall on one or other of the pitches of the chord.

HARMONICITY

The property of having a harmonic spectrum.

HARMONICS 4

The partials of a sound of definite pitch are (usually) exact multiples of some (audible) frequency known as the fundamental. In this case the partials are known as the harmonics of the sound.

HARMONISER 38

Application of brassage with tape-speed transposition to change the pitch of a sound without altering its duration. Generic name of commercially available hardware units which do both this and a number of other grain time-frame brassage procedures (e.g. duration change without pitch-change).

HARMONY

In European music, the rules governing the sounding-together of pitches, and the sequencing of chords.

HOMOPHONIC

Music in which there are several parts (instruments or voices) but all parts sound simultaneously (though not necessarily with the same pitch) at each musical event.

HPITCH

Pitch in a sense which refers to a HArmonic Field or to European notions of Harmony. In contrast to pitch as a property of a spectrum.

***** I *****

INBETWEENING 46

INDIAN RAG SYSTEM

The system of scales and associated figures and ornaments at the base of Classical Indian music theory and practice.

INHARMONIC

A spectrum which is not harmonic, but which is not noise, is said to be inharmonic. Inharmonic spectra may be bell-like (suggesting several pitches) or drum-like (being focused in some sense but lacking definitive pitch).

INTERPOLATION

The process of moving gradually between two defined states. See SPECTRAL INTERPOLATION.

INVERSE FOURIER TRANSFORM

A mathematical procedure which allows us to convert an arbitrary spectrum (frequency-domain representation of a sound) into the corresponding waveform (time-domain representation of the sound). See also, FOURIER TRANSFORM.

INVERSE KARPLUS STRONG

See SOUND PLUCKING.

ITERATION 42

***** J *****

JITTER

Tiny random fluctuations in apparently perceptually stable properties of a sound, such as pitch or loudness.

***** K *****

KARPLUS STRONG

An efficient algorithm for generating plucked-string sounds.

KEY

A piece of european music using the tempered scale (except where it is intentionally atonal) can usually be related to a scale beginning on a particular pitch, around which the melodic patterns and chord progressions of the piece are organised. The pitch which begins the scale defines the Key of the piece.

KLANGFARBENMELODIE

Musical line where successive pitches are played by different (groups of) instruments. Literally, tone-colour melody.

***** L *****

LOUDNESS

Technically speaking, loudness is a property which is related to perception, and is measured in a way which takes into account the varying sensitivity of the ear over different frequency ranges. In this sense it differs from the Amplitude of the signal, which is a scientific measure of the strength of a sound. In this book, the term loudness is used in the text, wherever this will not cause any confusion. Diagrams usually refer to Amplitude.

***** M *****

MAJOR

Most european music uses one of two scale patterns, known as the major scale and the minor scale (the latter having a number of variants).

META-INSTRUMENT

An instrument which provides control instructions for another instrument. E.g. the mixing of sounds is controlled by a mixing score (which may be a graphic representation of sounds and their entry times, or a written list in a computer file). A meta-instrument might write or modify the mixing instructions according to criteria supplied by the composer, or in response to other data.

MF

mezzo forte = moderately loud.

MIDI

Acronym for Musical Instrument Digital Interface. This is a communication protocol for messages sent between different digital musical instruments and computers. MIDI stores information on which key is pressed or released, how forcefully (or quickly) it is pressed, and on certain kinds of control information provided by controllers on digital instruments (e.g. pitch-glide, tremolo etc), together with more instrument specific data (which synthesis patch is being used). MIDI does NOT record the sound itself.

MINOR

Most european music uses one of two scale patterns, known as the major scale and the minor scale. The minor scale has two important variants, the melodic and the harmonic minor.

MIXING SCORE

A set of instructions detailing what will happen when a number of sounds are MIXED together. This might be a text file or a graphic display on a computer, but could equally well be a set of instructions for moving faders on a mixing desk in a studio. A typical mixing score would contain information about which sounds were to be used, at what time each would start, how loud each would be, and at what spatial position each should be placed.

MONO

Sound emanating from a single source (e.g. a single loudspeaker) or a single channel of digital information. As opposed to stereo.

MORPHOLOGY

The way in which the properties of a sound vary with time.

MOTIF

Small element of musical structure, usually consisting of a sequence of a few pitches (in notated music), and out of which larger structural units (e.g. phrases) are built.

MULTI-DIMENSIONAL SPACE

A line defines a one-dimensional space, a sheet of paper or the surface (only) of a sphere a two-dimensional space, and the world we live in is a three-dimensional space. We may generalise the notion of a space to any group of independent parameters. e.g. pitch and duration may be thought of as defining a two-dimensional space, and this is the space that we draw in when we notate a traditional musical score. Spaces may be of any number of dimensions (i.e. not necessarily ones that we can visualise in our own spatial experience) from the four dimensions of Einstein's space-time, to the infinite number of dimensions in Hilbert space.

MULTI-SOURCE BRASSAGE 45

***** N *****

NOISE

Sound having no perceptible pitch(es) and in which energy is distributed densely and randomly over the spectrum and/or in a way which varies randomly with time. Typical examples might be the consonants 's' or 'sh'. Other sounds (especially those recorded directly from the natural environment) may contain elements of unwanted noise which we may wish to eliminate by noise reduction.

NOISE REDUCTION

Process of eliminating or reducing unwanted noise in a sound source.

NOTCH FILTER 7

***** O *****

OCTAVE

A sound whose pitch is an octave higher than a second sound, usually has twice the frequency of that second sound. Two pitches separated by an octave, in european music, are regarded in some sense as the 'same' pitch, or as belonging to the same 'pitch-class'.

OCTAVE-STACKING 48

ONSET SYNCHRONISATION 48

***** P *****

PARAMETER

Any property of a sound or a sequence of sounds which can be musically organised, Parameter often also implies the measurability of that property.

PARTIAL 3

The sinusoidal elements which define the spectrum of a sound.

PARTIAL TRACKING 21

PERMUTATION

Specific rearrangement of the elements, or the properties of the elements (e.g. loudness), of a sequence of musical events.

PHASE 9

PHASE INVERSION .... (9)

The sound waveform may be replaced by the same form but 'upside-down' (i.e. the new wave rises where the other falls and vice versa). The same-effect is achieved with a sinusoidal waveform by restarting the wave from the first point at which it recrosses the zero-line (known as a phase-shift by 180 degrees, or by PI radians). Superimposing the original sound on its phase-inverted version causes the sounds to cancel one another.

PHASE VOCODER 11

Instrument producing a moment-by-moment analysis of a sound waveform to reveal its time-evolving spectrum. A windowed Fast Fourier Transform.

PHASING 9

PHONEME

A fundamental sound unit of a language, Crudely speaking we may think of vowels and consonants (as heard, rather than as written) but the true definition is more subtle.

PHRASE

Element of musical structure consisting of a sequence of sound events and, in traditional practice, usually lasting for a few bars.

PHYSICALITY

The physical nature of the apparent source of the sound. NOT the physical nature of the real source. Is the apparent source hard or soft, rigid or flexible, granular of of-a-piece etc.

PITCH... 4

A property of instrumental and vocal sounds organised in most traditional musical practices. Pitch arises from a regular arrangement of partials in the spectrum. All partials are multiples of some fundamental frequency which is audible (and which may or may not be present in the spectrum), and in this case they are known as the harmonics of the sound. In the simplest case (the sine wave) there is only the fundamental present. Humans hear pitch between approximately 16 cycles per second and 4000 cycles per second. Below 16 cps, the sound breaks up into a grain-stream. Above 4000 cycles we may still be aware of tessitura (relative pitch-range) but assigning specific pitch becomes more problematic.

PITCH-GLIDE

See PORTAMENTO.

PITCH-TRACKING 70,71

Finding and recording the (time-varying) pitch of a sound.

PITCH-TRACKING BY AUTO-CORRELATION 70

PITCH-TRACKING BY PARTIAL ANALYSIS 71

PITCH-TRANSFER

Imposing (time-varying) pitch of a sound on a different sound.

PORTAMENTO

Sliding of pitch. Often incorrectly referred to as glissando, but there is an important distinction. On a fretted or keyed instrument like a piano, we may slide fingers from a high pitch to a low pitch, but the keys allow us to access only the pitches of the scale, and we hear a rapid descending scale passage: a glissando. On a trombone or violin, a similar motion with the slide, or the finger along the string, causes pitch to fall through the continuum, without picking out intervening scale pitches: a portamento.

PROCESS-FOCUSSED TRANSFORMATION

Some musical processes will so radically alter a source sound that the perceived goal-sound is more dependent on the process of transformation than on the source itself. The same process applied to two very different sources will produce very similar goal-sounds. The perceived result of the procedure is governed more by the artefacts of the process of transformation, rather than by the particular nature of the source-sounds employed.

***** Q *****

Q 8

The steepness with which a filter cuts out unwanted frequencies.

QUANTISATION

Forcing the timing of events on to a time grid of a specific size. E.g. we may set a grid at 0.01 seconds. Any event must then fall at some multiple of 0.01 seconds. ALternatively we may set a grid at some division of the metrical unit e.g. a grid of demi-semi-quavers. All events must then fall on some multiple of demi-semiquaver divisions of the beat. The actual time quantisation will then also be determined by the tempo (the number of crotchets per second). The quantisation grid provides a time reference-frame. On keyed or fretted instruments, pitch is similarly quantised.

QUARTER TONE

A very small division of the musical scale. Half a semitone (the smallest interval accessible on a standard European keyboard instrument, like a piano).

***** R *****

RANDOM–CUTTING 41

REFERENCE FRAME

A set of values which provide a reference set against which other values can be measured. e.g. the chromatic scale as a reference set for european harmony, the set of vowels in standard english as a reference set for classifying regional accents etc.

RESONANCE

If an object is vibrated it will produce a sound. Due to its particular weight, size and shape there will be certain frequencies at which it will vibrate 'naturally'. If supplied with frequency–unspecific energy it will tend to vibrate at these natural resonant frequencies. A flute tube with a certain combination of closed holes has specific resonant frequencies which produce the pitches for that fingering. A hall or building will reinforce certain frequencies in a voice, orchestra etc which fall on its natural resonant frequencies.

RETROGRADE

The performance of a sequence of sound events in the reverse order. A–B–C–D–E becomes E–D–C–B–A. Note that the sound–events themselves are not reversed.

REVERBERATION 64

RITARDANDO

A decrease in the speed at which musical events succeed one another.

***** S *****

SAMPLER

A piece of hardware, or a software package on a computer, which that digitally records any sound and allows it to be manipulated (e.g. pitch–change by 'tape–speed' variation with the specific transposition information sent from a MIDI keyboard). The sounds recorded on a sampler are often referred to in the commercial literature as 'samples'. These should not be confused with the individual numbers used to record the shape of the waveform itself, which are properly known as samples. (See SAMPLING).

SAMPLING 1

Sound is digitally recorded by sampling the value of the (electrical analogue of the) sound wave, at regular time– intervals. These time intervals must be very short if high frequencies in the sound are to be resolved. (e.g. between 22000 and 48000 samples per second). At a sampling rate of 48000 samples per second, the highest resolvable frequency is 24000 cycles per second.

SCORE

The notation of a piece of music from which a performance of the work is recreated.

SEMITONE

The smallest interval between consecutive pitches on a modern European keyboard or fretted instrument. Musical scales are defined as some pattern of tones (equal to two semitones) and semitones, HArmonic minor scales also containing a three semitone step.

SEQUENCE GENERATION 1

SEQUENCES

Groups of consecutive sounds having distinctly different spectral properties. e.g. speech, or melodic phrases on keyed instruments (the spectra of whose elements differ by perceptually significant pitch steps).

SERIAL COMPOSITION

Style of twentieth century European musical composition in which the 12 pitches of the chromatic scale are arranged in a specific order, and this sequence (and certain well–specified transformations of it) are used as the basis for the organisation of pitches in a piece. The general idea of serialism was also extended to sequences of durations, of loudnesses etc. The even more general notion of permuting a given set of elements has been more widely used (e.g. systems music).

SERIALISM

See SERIAL COMPOSITION

SHAWM

A mediaeval wind instrument with a strong reedy sound.

SHEPARD TONES 72

Sounds (or sequences of sounds) constructed so that they rise in tessitura while their pitch falls (or vice versa).

SINE WAVE 2,5

The elementary oscillations in terms of which all other regular oscillations, vibrations or waveforms can be described. The oscillation of a simple (idealised) pendulum is described by a sine wave.

SINUSOIDAL

Having the shape of a sine–wave.

SOUND REVERSING 43

SOUND SHREDDING 41

SOUND-PLUCKING 72

A musical transformation which imposes a plucked-string-like attack on a sound.

SOURCE-FOCUSSED TRANSFORMATION

A musical transformation whose outcome depends strongly on the nature of the source sound. Defined in contrast to PROCESS-FOCUSSED TRANSFORMATION.

SPECTRAL ARPEGGIATION 24

SPECTRAL BLURRING 26

SPECTRAL BRIGHTNESS

A measure of where energy is focused in the spectrum. If some of upper partials are very loud, the sound will appear bright.

SPECTRAL FOCUSING 20

SPECTRAL FORMANT TRANSFER

see VOCODING.

SPECTRAL FREEZING 22

SPECTRAL INTERLEAVING 35

SPECTRAL INTERPOLATION 32-33

SPECTRAL MANIPULATION 18-35

Musical processes that work directly on the (time-varying) spectrum of the sound.

SPECTRAL MASKING 34

SPECTRAL SHAKING 23

SPECTRAL SHIFTING 18

SPECTRAL SPLITTING 29

SPECTRAL STRETCHING 19

SPECTRAL TIME-SHRINKING 30-31

SPECTRAL TIME-STRETCHING 30-31

SPECTRAL TIME-WARPING 30-31

SPECTRAL TRACE-AND-BLUR 27

SPECTRAL TRACING 25

SPECTRAL UNDULATION 28

SPECTRUM 2,3

Representation of a sound in terms of the frequencies of its partials (those sinusoidal waves which can be summed to produce the actual waveform of the sound). The frequency- domain representation of the sound.

SPLICING 40

The tail to head joining of two sounds. In the classical tape studio this would be achieved by joining the end of one sound tape to the beginning of another, using sticky tape.

SQUARE WAVE 5

SRUTI

The smallest unit into which the octave is divided in classical Indian music and from which the various rag scales can be derived. It is at least 6 times smaller than a semitone. This is more of a theoretical unit of measurement than a practically applied unit. In contrast, the Western semitone is built into the structure of its keyboard instruments.

STATIC INTERPOLATION

The process whereby a sound gradually changes into a different (kind of) sound during a series of repetitions of the sound, where each repeated unit is changed slightly away from the previous one and towards the goal sound.

STEREO

Sound emanating from two channels (e.g. two loudspeaker) or stored as two channels of digital information. As opposed to mono (from a single source). Sound information provided through two loudspeakers is able to convey information about the (apparent) positioning of sound sources in the intervening space between the loudspeakers, rather than suggesting merely a pair of sound sources.

STOCHASTIC PROCESSES

A process in which the probabilities of proceeding from one state, or set of states, to another, is defined. The temporal evolution of the process is therefore governed by a kind of weighted randomness, which can be chosen to give anything from an entirely determined outcome, to an entirely unpredictable one.

SYNTHESIS

Process of generating a sound from digital data, or from the parameters supplied to an electrical oscillator. Originally synthetic sounds were recognisably such, but now it is possible, through a process of careful analysis and subtle transformation, to recreate a recorded sound in a changed form which however sound as convincingly 'natural' as the recording of the original sound.

***** T *****

TEMPERED TUNING

The tuning of the scales used in European music., a system which became firmly established in the early eighteenth century. In tempered tuning the octave is divided into 12 exactly equal semitones. i.e. the ratio between the frequencies of any two pitches which are a semitone apart is exactly the same. In a harmonic spectrum, the frequencies of the partials are exact multiples of some fundamental frequency. The ratio of their frequencies form a pattern known as the harmonic series i.e.

2/1 3/2 4/3 5/4 6/5 7/6.....

and the frequency ratios between members of this series are known as 'pure' intervals. Some pure interval ratios are...

octave..2/1 4th.....4/3
5th.....3/2 7th.....7/4

There is no common smaller interval from which all these 'pure' intervals can be constructed. Hence the striving to achieve some kinds of compromise tunings, of which the tempered scale is just one example. Apart from the octave, the tempered scale only approximates frequency ratios of the pure intervals. The 7th has no close approximation in the tempered scale.

TEMPO

The rate at which musical events occur. In european music the relative duration of events are indicated by note values e.g. a crotchet is as long as two quavers. The speed of the whole will be indicated by a tempo marking e.g.

crotchet = 120

which means there are to be 120 crotchets in one minute.

TESSITURA

Also known as register. In its simplest sense, the range of pitch in question. The tessitura of soprano voices is higher than that of Tenor voices. However, unpitched (e.g. noise) sounds may rise in tessitura while having no perceivable pitch, while SHEPARD TONES may rise in tessitura while falling in pitch(!). Hence tessitura also has something to do with where the main focus of energy lies in the spectrum (where the loudest groups of partials are to be found).

TEXTURE

Organisation of sound elements in terms of (temporal) density and field properties.

Texture in which small sets of sound elements are considered as the organisable units of the texture, rather than individual sounds themselves. The sounds forming any particular group are, however, chosen arbitrarily from the available sound sources.

Texture in which specific small sets of sound elements, known as motifs, are considered as the organisable units of the texture, rather than individual sounds themselves. The sounds forming any particular motif are in some kind of predefined arrangement (e.g. some combination of sequence, time-placement, pitch-relationship, formant sequence, etc).

Texture in which specific small sets of sound elements, known as ornaments, are attached to the fundamental sound elements of the texture, and each of these may be organised as units of the texture.

TEXTURE STREAM

Dense and relatively disordered sequence of sound events in which specific ordering properties between individual elements cannot be perceived. Properties may be ordered in terms of Field and Density.

THRESHOLD

A value either not to be exceeded (or fallen below), or to be noted (in order to cause something else to happen) if it is thus exceeded.

TIMBRE

A catch-all term for all those aspects of a sound not included in pitch and duration. Of no value to the sound composer!

TIME STRETCHING

See spectral time-stretching, granular time-stretching, waveset time-stretching, harmoniser, brassage.

TIME-DOMAIN REPRESENTATION

The waveform of a sound represented as the variation of air- pressure with time.

TIME-VARIABLE TIME-STRETCHING

See time-warping.

TIME-WARPING 30

Changing the duration of a sound in a way which itself varies with time.

TONAL MUSIC

Music organised around keys and the progression between different keys. In contrast, atonal music avoids indicating the dominance of any particular key or pitch.

TONE

The interval between the first two pitches of a major (or minor) scale in European music. European scales consist of patterns of tones and semitones (half a tone) and, in some cases 3-semitone steps.

TRAJECTORY

The variation of some property with time. e.g. loudness trajectory, pitch trajectory, formant trajectory. Loudness trajectory is often also known as 'envelope' and instruments which manipulate the loudness trajectory are here called 'Envelope something'.

TREMOLO 66

Cyclical undulations of loudness between c. 4 and 20 cycles per second.

TRANSPOSITION

Changing the pitch of a sound, or sound sequence.

TRIGGERING 62

Using the value of some time-varying property (usually the loudness) of a sound to cause something else to happen.

TRITONE

The musical interval of 6 semitones. In the European tempered scale, a frequency ratio of the square root of 2, to 1.

***** U *****

UNDULATION

cyclical change in the value of a musical property, like the rising and falling of pitch in vibrato, or of loudness in tremolo.

***** V *****

VIBRAPHONE

A musical instrument consisting of pitched metal bars suspended over resonating tubes. When the bars are struck they produce a specific pitch. and the associated tube resonates at that pitch. A motor drives a small rotary blade inside the tube, adding slight vibrato-tremolo to the resonated sound.

VIBRATO 66

Cyclical undulations of pitch between c. 4 and 20 cycles per second.

VOCODING 34

***** W *****

WAVELENGTH 3,4

WAVESET 50

WAVESET DISTORTION 52

Generally refers to any process which irreversibly alters the wavesets in a sound e.g. waveset inversion, omission, reversal, shaking, shuffling, substitution, averaging and harmonic distortion. Specifically used to refer to power-distortion i.e. raising each sample value of the sound to a power (e.g. squaring, cubing, taking the square root).

WHITE-OUT

A musical process in which a texture becomes so dense that all spectral detail is lost, producing a noise band.

WINDOW

In sound analysis (conversion from waveform to spectrum, to spectral envelope or to time-varying filter-coefficients : see LPC) a window is a very brief slice of time (a few milliseconds) in which we make a spectral analysis, before passing to the next window to make our next analysis. We hence discover how the spectrum changes in time. Not to be confused with analysis CHANNEL.

WINDOWED FFT

A Fast Fourier Transform that is performed over a brief slice of time (a window), then performed over and over again at successive windows throughout the entire duration of the sound. The Phase Vocoder is a windowed FFT.

***** Z *****

ZERO-CROSSING

The waveform of a sound continually rises above the centre line and then falls below it. The point where it crosses the centre line is a zero-crossing.