AUDIBLE DESIGN

A PLAIN AND EASY INTRODUCTION TO PRACTICAL SOUND COMPOSITION

by

TREVOR WISHART

Published by Orpheus the Pantomime Ltd : 1994

Copyright: Trevor Wishart: 1994
BRIEF BIBLIOGRAPHY

DIE REHE
Journal, published by serial composers associated with the Darmstadt Summer School in the 1960s.

GESANG DER JUNGLINGE
Electronic acoustic work by Karlheinz Stockhausen, using boy’s voice and electronics. Recording available.

GRUPPEN
Large scale work for 3 orchestras surrounding the audience, by Karlheinz Stockhausen. Recording available.

ON SONIC ART

PITHOPRAKTA
Work by Xenakis Xenakis for orchestra, exploiting means of portamento and using statistical formulae to control textures.

RED BIRD

VOX 3
Work for 4 amplified voices by Trevor Wishart, from 1987, in which the voices sing in polyphony, coordinated by the use of independent but computer-generated click-tracks. Recording available.

VOX 5
Electro-acoustic work by Trevor Wishart, from 1986, in which a human voice is transformed into the sounds of many other recognisable sounds. Recording available.
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, so 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 test 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.

# 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 to compose in residence by Chasine Barlow, in 1993. Some classifications and supplementary material was added after discussions with Milker Puckette, Zark Semel, Stephan Bilbao and Philippe Deprato at IRCAM. However, the basis for any inaccuracies or omissions in the exposition rests entirely with me. The text of the book was originally written entirely in longhand and I am indebted to Wray Calmany for transcribing these personal typographical errors into computer files. Help with the final layout was provided by Tony Myan of the University of York.

My career in music-making with computers would not have been possible without the assistance 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 Composer Desktop Project and would like to thank my fellow contributors to this long running project; in particular Tom Findlay, the real driving force at the heart of the CDP, to whom we owe the survival, expansion and development of this collaborative venture, against all odds, and whose persistence proving and questioning has led to clear instrument descriptions & musicians’-friendly documentation. Richard Oron and Andrew Brandon, the other composer supervisor members of, and core instrument contributors to the project; David Mulhern who devised the hardware basis of the CDP and taught me, 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 closely learnt to program in anaehistically, if not elegantly); Rajpal Fieldman of the University of Keele, who has been principally responsible for developing the various graphic interfaces to the system; and to Michael Clarke, Nick Lavers, 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 VOR Cycle or Tangos of Fire, except for the one, Twenty, and I would like to thank Paul de Manis 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 Newson, for assistance with data transfer to DAT.
AUDIBLE DESIGN

A PLAIN AND EASY INTRODUCTION
TO PRACTICAL SOUND COMPOSITION

by

TREVOR WISHART

Published by Orpheus the Pantomime Ltd: 1994

Copyright: Trevor Wishart: 1994
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, wavelet time-stretching, wavelet time-stretching, harmonics, transients.

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 organized around keys and the progression between different keys. In contrast, non-tom 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, 5-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".

TREMOLI ................................. 66

Cyclical undulation 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.
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.

ZERO-CUTTING ........................................ 40
ZIGZAGGING ........................................... 43
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 deterministic 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 recognizably such, but now it is possible, through a process of careful analysis and subtle transformation, to increase a recorded sound in a changed form which however sound as convincingly 'natural' as the recording of the original sound.

TAPE-ACCELERATION .................. 36

TAPE-SPEED VARIATION ................. 36

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 ratios of these frequencies form a pattern known as the harmonic series i.e.

\[ 2/1 \ 3/2 \ 4/3 \ 5/4 \ 6/5 \ 7/6 \ldots \]

and the frequency ratios between members of this series are known as 'pure' intervals. Some pure interval ratios are...

octave: 2/1
5th: \( 3/2 \)
7th: \( 4/3 \)

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 is 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.

\[ \text{crotchet} = 120 \]

which means there are to be 120 crotchets in one minute.

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 Varèse was imagining a third unstable music which had the same degree of control over sonic substance as musicians have traditionally exercised over melody, harmony and rhythm. This concern grew directly out of the sophisticated development of mechanization in the late Nineteenth Century and to immense 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 unorthodoxized tools (in the early days, clicking between larger disks – later the transformations – like tape speed variations, editing and mixing – available with magnetic tape) masterpieces of this new medium began to emerge and an approach to musical composition sounds 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 ciphers certain well defined processes of transformation may have on these. We will also present, in the Appendices, a simple diagnostical 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 addressed towards the end of this book but in this detail as to date, no universal tradition of large scale form-building through these newly audible 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 Computers Desktop Project (CDP) System at the time of writing. The CDP was developed as a composer's cooperative and originated in York, U.K.

May we advise this chapter, discuss whether or not any of the processes described can be, or ought to be, implemented in real time. In other words, many of them will run in real-time environments. My concern here, however, is to explore the musical possibilities and contrasts offered by the medium of music 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 assign and transport 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 deeply available technology is based on the movement of sound. To a certain extent, 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/lyric 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 in 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 emphasize strongly enough that my concern is with the world of sound itself, as opposed to the world of emotions of sound, or the largely literary disciplines of music score analysis and criticism. It will soon be on what can be audibly perceived, on my direct response to these perceptions and on what can be technically, scientifically or mathematically described.

SPECTRAL TIME—SHRINKING 30—31
SPECTRAL TIME—STRETCHING 30—31
SPECTRAL TIME—WARPING 30—31
SPECTRAL TRACE—AND—BLUE 27
SPECTRAL TRACING 25
SPECTRAL UNDULATION 28
SPECTRUM 2.3

Representations of a sound in terms of the frequencies of its partials (those sinusoidal waves which can be summed to produce the sound waveform of the sound). The frequency—domain representation of the sound.

SQUEEZEING 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

SUITE

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 loudspeakers) or sound 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.
A musical transformation which imposes a plucked-string-like attack on a sound.

SOURCE-FOCUSED TRANSFORMATION

A musical transformation whose outcome depends strongly on the nature of the source sound. Defined in contrast to PROCESS-FOCUSED TRANSFORMATION.

SPECTRAL ARPEGGATION .......... 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 SHADING ............. 18
SPECTRAL SPLITTING ............. 29
SPECTRAL STRETCHING .......... 19

The world of sound-composition has been hampered by being cast in the role of a pure 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 stretch into existence against the immense inertia of received wisdom about "musical structure".

I have discussed elsewhere [On Sound Art] the strong influence of intellectual "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 intrinsic to the scientific method and to technological advance.

I don't wish here to decay the idea that there may be "universal truths" about human behaviour and human social interaction which scientific and technology are powerless to alter. But because our prime medium for the propagation of knowledge in 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 inappropriately.

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 potential 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 accept 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 option to be decided on the authority of a musical text (a score or a book... even this one), or the importance of the scholarly 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 not merely not texts like this book, or even true scientific theories, but computer programs themselves. Here, the religious or mystical potency with which the musical text was once 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 subvert sonic experience. Such texts are potent but, at the same time, judgemental. They do not radiate some mystical authority to which we must bow, but do something specific in the world which we can judge to be more or less successful. And if we are disillusioned, 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 result of this process (as well as a safeguard against losing my heart-renowned public) I would keep copious notes and copies of many of the intermediate stages or make reconstruction (or variation) of the final sound possible. Today, I store only the "core" and a brief so-called "batch-file." The latter is the "core," if activated, will automatically run all the programs necessary to create the final sound. It can also be copied and modified to produce whatever variations are required.

An illuminating maneuver 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 these 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 in past musical craft. The musical world is generally conservative and denizens of this community can be quick to dismiss the "new-fangled" as semiotically 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 algorithms. In many cases, all the processing in sound-processing as it applies to sound signals. However, the motivation of our discussion is somewhat different from that of the scientific or technological presentation of signals. In analyzing and transforming signals for scientific purposes, we normally have some distinct goal in mind—i.e., 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 merit of our procedures against the desired outcome and hence assess the validity of our procedure.

In some cases, musicians share these goals. Precision analysis of sounds, extraction of time-varying information in the frequency domain (see Appendix p. 97), sound clarification or noise reduction, are all of great importance to the sonic composer. But beyond this, the question that a composer asks is, if this process aesthetically interesting?—does the sound resulting from the process reflect perceptually and is it artistically useful—that is how we begin? What are we looking for in a way to transform sound material into giving results which are already close relatives of the sources, but also clearly different. Ideally we require a way to "add-on," or order three degrees of difference allowing us to articulate the space of sound possibilities in a structured and meaningful way.

Beyond this, there are no constraints restrictions on what we do. In particular, the goals of the process we act in motion may not be known or even (with complex signals) easily predictable beforehand. In fact, as required, 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 methodologies of the scientific task may involve the design of a high-pass filter to achieve a particular result. A musician, however, is more likely to require an exceptionally flexible (e.g., variable, time-variable, Appendix p. 97) filter in order to explore its effects on sound materials. He/herself may not know beforehand exactly what is being searched for when it is used, apart from the useful aesthetic transformation of the original source. When this means to practice may only emerge in the course of the exploration.

**SCORE**

The notation of a piece of music from which a performance of the work is recreated.

**SEMITUONE**

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**

**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-specifed transformations of it) are used as the basis for the organization of music in a piece. The general idea of serialism was also extended to sequences of durations, of dynamic, etc. The even more general notion of permuting a given set of elements has been more widely used (e.g., systems music).

**SHAWM**

A medieval wind instrument with a strong reedy sound.

**SHEPHERD TONES**

Sounds (or sequences of sounds) constructed so that they rise in treble while their pitch falls (vice versa).

**SINE WAVE**

The elementary oscillations in terms of which all other regular oscillations, vibrations or waveforms can be described. The oscillation of a simple (idealized) pendulum is described by a sine wave.

**SINUSODIAL**

Having the shape of a sine-wave.
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-specific 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 fingerering. A hall of 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.

SAMPLER

A piece of hardware or a software package on a computer, which can digitally record any sound and allows it to be manipulated (e.g. pitch-change by 'tape-speed' variation with the specific transportation information sent to 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 effective amplitude of the sound wave, at regular time-intervals. These time intervals must be very short if high frequencies in the sound are to be recorded. (e.g. between 2000 and 48000 samples per seocnd). At a sampling rate of 48000 samples per second, the highest receivavle frequency is 24000 cycles per second.

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, noise-crunching program will produce an arbitrary result - like the old adage of trying to write a play by tossing a chimpanzee maw away at a typewriter in the hope that a masterpiece will emerge.

So science be forewarned! We may embark on signal-processing procedures which will appear haimonous to the scientifically sophisticated, procedures which give relatively ungodly results, or that are heavily dependent on the unique properties of the particular signal 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 maker. Much has Twenty Common Westeners music been based on an obsession with complications. This has arisen partly from the pernicious intransigence of late romantic and also from an intellectually suspect linkage of form to information theory with musical communication - a notion which means more information, means more musical 'power'. This obsession with quantity, or information overload, arises partly from the breakdown of consensus on the substance of musical meaning. In the end, the core of musical potency remains as elusive as ever but in an age which demands everything be quantifiable and measurable, a model which renews the quantity of information, or complications of an artifact, seems hardly plausible.

This demand of overload 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 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 ear and this feedback is used to modify the action. Learned through physical practice and repetition of others and aided by discussion and chanting 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 utterances 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 content of a particular vocal 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 targeted 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
detailed mock (the score) which can often be given a rational em em ph em ph em ph em ph.

Some composers (particularly the discarnate "romantic" composers), may use the musical score merely as a receptacle for their sensitive 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 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 is required to do, suggesting therefore that
the computer must also have a clearly explicit description of the task. I will describe here why this is
not so.

The absurd and misguided rationalist dogma of the totally explicit construction of all aspects of a
musical work is not the "natural" outcom of the use of computer technology in musical composition

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 composer and
composer role. The fact, however, is that most electro-acoustic composers play with the medium, playing
through insulated 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 "synthesiser is a taking place between
composer and performer activities and as the activity of composer and performer begins 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
has responsibility for the organisation of certain well-controlled parameters (pitch, duration,
intensity 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, tuning
precision or its interpretation, and sound production and articulation.

At the other pole, the free-improvising performer relies on an intuitive creative process (influenced by
the intrinsic sound/pitch limitations of a sound source, e.g. a resaun piano, a metal sheet) to generate
both the moments-continuous 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, the instrument builder
has a regulating role similar to that of the composer.

PORTAMENTO

Siding of pitch. Often incorrectly referred to as a glissando, but there is an important distinction. On a
flute or oboe or clarinet 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

PROCESS-FOCUSED 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 process is
governed more by the audibility of the process of transformation, rather than by the particular nature
of the source-sounds employed.

QUANTISATION

Forcing the timing of events as to a time grid of a specific size. E.g. we may set a grid at 01.00 seconds.
Any event then falls at some multiple of 01.00 seconds. Alternatively we may set a grid at some
division of the instanta�数 unit e.g. a grid of dots-secs-quavers. All events must then fall on some
multiple of dots-secs-quavers divisions of the beat. The actual time quantisation will then also be
determined by the tempo (the number of crotchetts per second). The quantisation grid provides a time
reference-frame. On keyed or formed 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).
PHASE VOCODER


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. In the apparent source hard, soft, rigid or flexible, granular or of-a-piece etc.

ITALIC

A property of instrumental and vocal sounds organized 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 a fundamental frequency. Humans hear pitch between approximately 16 cycles per second and 4000 cycles per second. Below 16 cps, the sound breaks up into a noise-spectrum. Above 4000 cycles we may still be aware of texture (relative pitch-range) but assigning specific pitch becomes more problematic.

PHYSICALITY

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

Learning (time-varying) pitch of a sound on a different sound.

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 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 (or at least configurations) for a particular purpose. The "instrument" is no longer a definable (if subtle) clumped universe but a groundswell of possibilities out of which the sonic composer must define some semantically valid universe. Without some kind of intuitive understanding of the inverse of sounds, the problem of choice is innumerable (unless one replaces it entirely by non-choice strategies, dice-throwing procedures etc).

REAL TIME OR NOT REAL TIME?

That is the question

As composers become faster and more, more and more powerful, there is a cluster among musicians working in the medium for "real-time" systems, i.e. computer 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 strike a violin, we immediately hear a sound being produced which is theresult of our decision to use a particular finger position and how pressure and which responds immediately to our subtle manual articulation of these. Success in performance also depends on a frontknowledge of what our actions will precipitate. Hence the rationale for having real-time systems seems quite clear in the sphere of instrumental performance. To develop new, or extended musical performance instruments (including real-time computer or real-time 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, into a recording medium not 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 "instant" (the theater versus the cinema, the work is recreated before your very eyes) and of insatiable dependent on each performer's interpretation of the work. This approach fits well into the 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 recognize that he/she is in fact creating a new and different instrument for the instrumentalist (or instrumentalk-technician- duc) to play and in partly adopting the role of an instrument builder with its own very different responsibilities. In the acoustic 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
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 composite music, are regarded to some sense as the "same" pitch, or as belonging to the same "pitch-class".

OCTAVE-STAKING  .......... 48

ONSET SYNCHRONISATION  .......... 48

PARAMETER

Any property of a sound or a sequence of sounds which can be musically organized. 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 which, when superimposed on the original produces the same as if it were inverted).
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 generalize 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 write a traditional musical score. Spaces may be of any number of dimensions (i.e. not necessarily ones that we can visualize 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 BRASSAGI  .......... 45

NOISE

Sound wrong no perceivable 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 't'. 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

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 sound

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

In this interface between intuitive knowledge embedded in bodily movement and direct audio feedback, and explicit knowledge carried in a numeric 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 encoded in performance practice) and the skill of the computer 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 rules 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 so very lack established limits at this moment, it may turn out to be intrinsically unbounded. The exploratory searching 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 art of sound-sources (instruments) and a deeply embedded performance practice, perhaps these supermen exist. In contrast, however, good sound composition always involves a process of discovery, and hence a coming to terms with the unexpected and the unreal, 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. Some Art is not like making.
We already know that sound composition presents us with many modes of musical—variation which are
unavoidable— at least in a specifically consumable 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.

MAJOR

Most European music uses one of two scales, 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

metzo forte = moderately loud.

MIDI

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 or digital instruments (e.g., pitch-bend, tempo etc.), together with more
Instrument specific data (which synthesizes pitch 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.

MIX—SHUFFLING ......................... 47
MIXING .................................. 46
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, or at what time each would start, how loud each would be, and in 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.
KAPPLER STRONG

An efficient algorithm for generating plucked-string sounds.

KEY

A piece of computer music using the tempered scale (except where it is intentionally a stroda) can usually be related to a scale beginning on a particular pitch, around which the melodic pattern and chord progressions of the piece are organized. The pitch which begins the scale defines the Key of the piece.

KLINGFARBNELMELONE

Musical line where successive pitches are played by different groups of instruments. Literally, tone-colour melody.

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 difference is black enough to serve my purpose. 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 abstractions. "E minor" is a form related to "E-flat major" in a chord. But these are relations between ideals, or classes, of sounds. For every "E minor" 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 abstractions properties of real sounds—the pitch, the duration, the loudness. Certain other features like pitch stability, melodic and vibrato control etc. are expected to be within perceptible limits defined by performance practice. But beyond this there is a less well-defined sphere known as interpretation. In this way, traditionally, we have a structure defined by relations among archetypical 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 immense consequences which force us to extend, or even to invent, traditional musical practice.

To give just 2 examples...

1. We can capture a very special articulation of sound that cannot be reproduced through performance—a particular note in which a phrase is spoken on a particular occasion by an unknown speaker; the resonant 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 arises out of the special combination of performers involved and the energy of the moment.

2. Exact reproduction allows 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... the exact repetition of sonic events at an exact time interval responds to a fixed frequency and hence a perceived pitch.

The most important thing to understand, however, is that a sound is a sound and 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 dust-grain), pitch or pitch-band, pitch motion, spectral harmonicity—inharmonicity and its evolution, spectral contour and formant (see Chapter 3) and their evolution, spectral stability and its evolution, and dynamic modulation and spectral continuons (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
computational 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 procedural 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 formalisms of notational conventions (including theoretical models relating notation and clavichord) with a pool of sound resources from which musical "buildings" could be constructed. Compositions through traditional instruments are used together as a clear 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 bowed strings (violin), and writing 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 safely assessed 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 select only 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 combining isometric buildings in which every pebble is of a
different color. As the Twentieth Century has progressed and the possibilities of conventional
instruments have been explored to their limit, we know how to recognize various shades of grey and
gold to make our 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 not – depending on the light. These classification categories are undermined and our original task seems to become overwrought. We need a new perspective to
understand this new world.

In sonic terms, not only sounds of transistors (like a simple percussion, definable portamento, or
inharmonic spectra) but those of unstable, or rapidly-varying spectra (the growing gas, the human
speech-sounds) must be accepted into the compositional universe. Most sounds simply do not fall into
the neat categories provided by a reduced–instrument oriented conception of musical architecture. As
most traditional building plan and pitch (and duration) as their primary ordering principles, working
with 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 combining 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

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 hard-like (suggesting several pitches) or dream-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 KARPPLUS STRONG
See SOUND PLUCKING.
ITERATION .............................................. 42
JITTER
Tiny random fluctuations is apparently perceptually enable properties of a sound, such as pitch or loudness.
HARMONIC

In traditional European practice, harmonic means pertaining to harmony and refers particularly to tonal music. In sound composition, harmonic refers to the property of a sound where all the partials are multiples of some (audible) fundamental frequency. These partials of a sound are known as harmonics. In this context, the sound is said to be harmonic and the sound has a natural, definite pitch. (See also INHARMONIC). These two terms are incompatible, so in the case we see Harmonic to refer to the traditional usage, and harmonics 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.

HARMONICIZER .................................................. 38

Application of processing 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 processing 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.

HINTCH

Pitch in a sense which refers to Harmonic Field or to European notions of Harmony. In contrast to pitch as a property of a spectrum.

numerical harmony, separate its constituents, merge the constituents from two quite different pedigrees 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 chemistry, we can take apart what was once the raw materials of music, recombine them, or transform them into new and un梦想 of musical materials. Sound becomes a fluid and entirely malleable medium, not a carefully honed collection of gongs. Sculpture and chemistry, rather than language or finite mathematics, become appropriate metaphors for what a composer might do, although mathematical and physical principles 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 incommensurable features of sounds may be analyzed, understood, transformed and transformed in rigorously definable ways. A minute audible difference of a particular sound can be magnified by time-stretching or breath into huge by cyclic repetition (as in the works of Steve Reich). The evolving spectrum of a complex source event can be pared away until only a few of the constituent partials remain, transforming something that was once smooth and jangled, into something remarkable (spectral scoring: see Chapter 3). We may exaggerate or contrive – in a precise manner – the energy (loudness) trajectory (or envelope) of a sound, reducing or contrasting its general propensities, and we can pass between these states of being of the sound with complete facility, tracing out an audible path of musical connections – a trace for musical forms – building.

This shift is emphasized as it 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 search, mould and transform this continuum in any way we choose, to build new worlds of musical connectedness. To go any further in this universe, we need to understand the properties of the "onic matrix" 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 observe 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 pushing 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 water where the air molecules move on water in one direction. It also explains how sound can travel very much faster than the most furious 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) distort 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 in alternately 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

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 wave.

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.
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 has 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

***** C *****

GATING ........................................... 60

GOLDEN SECTION

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

\[ AP \cdot BP = AB \]

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

\[ A = \frac{1}{2} \quad P = \frac{1}{2} \quad B \]

Diagram 2

Drum rotates
as shown.

Point pen drawn
on drum.

Aneroid barometer:

Aneroid barometer stack expands and contracts 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 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 aware 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 3).

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 essentially 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 soundwaves 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 existing 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) synthesizers.

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 recorded to time-varying 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 to-day one 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 fluctuations is converted into a pattern of numbers. (See Appendix E). The individual numbers are known as samples and are often referred to as "samples", and each one represents the instantaneous (instant) value of the pressure in the original air pressure wave.
ECHO .................................................. 64

EDITING

The process of editing sounds into shorter segments or splitting longer sounds into 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 ................. 59,90
ENVELOPE TRANSFORMATION ............. 50

Musical transformations of the loudness trajectory of a sound.

ENVOLVING .................................... 59
EXPANDING ................................. 60
We now arrive at an essentially abstract representation of sonic substance for, although these numbers are normally stored in a 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 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 notation practice. Traditional notation was an abstraction of general and easily quantifiable large-scale properties of sound events (or performance gesture, 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 metaphors 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 (in the Eighteenth Century French mathematician investigating heat distribution in solid objects) who realised that any particular wave—shape (or, mathematically, any function) can be expressed 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

---

**DETECTION**

Describes the way in which a range is filled. Applies particularly to time ranges. A high-density texture has a large 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, sonance etc), rather than the pitch or duration. Distortion also implies the degradation of the sound (and in variability means that the original sound cannot be recorded from the distorted version). Destructive distortion which preserves wave—shaping 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**

Mean of ensuring the prominence of a loud "voice" in a mix.

**DURATION**

The length of time a sound persists. Not to be confused with event—onset—separation—duration, 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.
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. braying, zigzagging).

CONSTRUCTIVE DISTORTION

A process which generates musically interesting artifacts 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 morphologies). These sounds have continuity.

CONTOUR

The shape of some property of a sound at one moment in time. In music, 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 brightness). To avoid any confusion, this book reserves Envelope for such audio properties, and Contour for the instantaneous shape of a property. The masses of commercial instruments may however use the term Envelope.

The overall 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.

COSOUND

A general purpose computer language which allows a composer to describe a sound-generating procedure (synthesis method) and ways to control it, to 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. Cosound 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
The first significant object from a musical point of view is a shape made up 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, beat, spectrum etc.) to the sound (for a fuller discussion of this point see On Sonic Art).

This perceptual boundary is neatly illustrated by the process of wavecycle time-averaging. This process lengthens a sound by repeating each wavecycle (a wavecycle is akin to a wavecycle but not exactly the same thing. For more details see Appendix page 55). With noisy sources, the waveforms vary widely from one to another but we hear only the result, a sound of indefinite pitch. Repeating each wavecycle lengthens the sound and introduces some artifacts. If we repeat each wavecycle 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 rigid stream of picked beats. (Sound example 1-9).

Once we can perceive distinctive qualitative characteristics in a sound, we have arrived. The boundary between the wavecycle time frame and the gram-time frame is of great importance in instrument design. For example, imagine we wished to separate the grains in a sound (like a roll of 7") by examining the loudest trajectory of the sound. Intuitively we can say that the grains are the loud part of the signal, and the points between grains the quieter parts. If we set 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 seek the instrument to do is search for a peak in the signal where the signal stays below the (absolute) gate value for a significant length of time. This time is at least of gram-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 content, onset characteristics (hard-edged, soft-edged), pitchy/hoarse/airy quality etc. The grain is characterized 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", "Y", "p", "s".

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 spikes as in a hi-hat), yet no particular pitch may be discernible. Analyses of such sounds may reveal either a very short infrasonic spectrum (which has insufficient time to register as several pitches, as we might hear in an inharmonic bell sound), or a rapidly changing pitched spectrum. Although the internal structure of each sound is the cause of what we hear, we do not resolve this internal structure in our perception. The experience of a grain is indivisible.

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, robbed, shaken, spoken etc.? ??

CHANGE RINGING
A form of bell-ringing in which the order in which the bells are sounded is permuted in specified 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 30 and 150, 150 and 250, 250 and 350 etc) or, more discriminately, 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
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 semitones divisions of the octave.

COMB-_FILTER TRANSPOSITION.... 65
A quick method of achieving octave-ward transposition of a sound, where its pitch is known.

COMB-FILTERING............ 64
BANDPASS FILTER

BAND SELECT FILTER

BAR

A grouping of musical duration-units. The bar-length is measured in some basic musical unit (e.g. crochets, quavers) for which a speed (tempo) is given (e.g. 120 crochets per minute). In most music, 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 replaces these together tail to head. In the simplest case sounds are selected in order, from the source, and reglued 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.:

<table>
<thead>
<tr>
<th>time (sec)</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>2.7</td>
</tr>
<tr>
<td>1.3</td>
<td>2.2</td>
</tr>
<tr>
<td>1.7</td>
<td>2.1</td>
</tr>
</tbody>
</table>

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.

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 indwelling was brought home to me through working with birdsong. A particular song consisted of a rapid-repeated sound having a "bulky" quality. One might presume from this that the individual sounds were internally parameterized at a rate too fast to be heard. In fact, when slowed down to fifth speed, each sound was found to be composed 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 timbre and hence the individuality and qualitative unity of a grain and we will return to this later. But if the sound remains beyond a new time limit (around 0.20 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 individual 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 Amplitude, we chop up a sound into time 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 too much so that the resulting sound, we will clearly end up with a longer sound (Appendix p.44).

If we try to make segments smaller than the grain-size, we will destroy the signal because the overlap (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 result in parts of the waveform inaudible, to make a waveform twice as long, and our sound will drop by an octave, as in speed-up velocity. If the segments are in the grain time-frame, the incommensurate pitch will be preserved intact, and we should time-stretch the source without changing its pitch (the harmonic algorithm). If the segments are longer than grains, their internal structure will be heard out and we will have 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 1.3).

The grain/continuation time-frame boundary is one 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 (cycle, 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) as 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-contraction into a structurally inseparable 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 (shape, width) of the spectrum, with the articulation of pitch (vibration, jets), loudness (transients), spectral contour or formant gliding (see Chapter 3) etc. etc. The continuation may, however, be discontinuous as in iterative sounds (grain-sequences - such as rifled "B", and low transients as in which are processed partly as a rapid sequence of grains) or segmented sounds (see below), or granular sentence streams (e.g. murmurs.

APPENDIX 1

(Page numbers refer to the diagrammatic Appendix 3).

... **** ACCELERANDO

As increase in the speed at which musical events occur one another.

ACUMULATION

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-sign. This normally remains intact in perception as the loudness of the sound, but amplitude and loudness are 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 generalized to include any process attempting to uncover the underlying structure of a piece of music.

ARTICULATION

The (consonant) 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.
with coarse pitch) where the sound is clearly made up of many individual overlapped attacks. Here, variation in grain or segment speed or density and grain or segment qualities will also contribute to our mental experience of continuations.

As our time-frame lengths, we reach the realm of the Phrase. Just as in traditional musical practice, the boundary between a long undertone and 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 context. A well-known and short, lasting over four bars may be regarded as a note-attenuation (an example of con-excitation) and may exceed in length a melodic phrase. But still with a marked series of changes and speed changes might well function as a musical phrase (depending on context). (Sound example 5:3).

A similar ambiguity applies in the sound domain with an added difficulty. Whereas it will usually be quite clear what is a new event and what is a phrase (meaningless and unifies the events), a sound event can be infinitely complex. We might, for example, start with a spoken sentence, a phrase—time-frame object, then time-shift 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 method will allow us to break it, 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 sound time-frames up to the level of the duration of an entire work. These involved time-frames are the basis of our perception of both rhythm and large-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 will be found in On Sonic Art and in the writings of the Group de Recherches Musicales. How, I would like to draw attention to two aspects of our musical experience.

The way in which a sound is attached 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 get an impression of an recognizable 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 assume that the goal sound is related to the source sound merely because we have derived one from the other. We have to pay a closer attention to the experience of the listener either through clear aural, morphological, or no similarities between the sounds, or a clear path through a sense of connecting sounds which gradually change their characteristics form sounds 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 sequentially time-extend and change the loudness trajectory (envelope) of a vocal sound, producing word-like attacks, which are then progressively damped to sound like unpitched drum sounds. (Sound example 1.4).

In general, sounds may be achieved 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 gracht – a xylophone note, a drum stroke, a vocal click. It may be short, but permitted to resinate through an associated physical system, e.g. a cello sounds out for a pianissimo note, a resonant bell for a drum stroke. Or the material itself may have internal resonating properties (halls, gongs, metal reeds) 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, oboe, violin. The medium may vibrate at a frequency related to the medium's form (e.g. water-droplet sizes, 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 interval sound (called "C", drum roll etc.).

The vibrating medium itself may be classically mobile – a flautophone, a fluted metal shank, a manuscript larynx – so that the pitch or spectrum of the sound varies through time. The material may be only gently caressed into motion (the air in a "Blowing", the strings in a Rammelklok) giving the sound a soft quality, or the medium may be loudly bowed and granular (the sand or beach in a Shaker or Wind-machine) giving the sound a diffuse continuity. Resonant systems will stabilize after a varying of transient or unstable initial events (flute bowings, acoustic guitar bows to a metal shank) so that a complex and disconnected onset leads to a stable or slowly evolving spectrum.

I am not suggesting that we consciously analyse our aural experience in this way. On the contrary, such analysis is so important as to be almost considered an intuitive (not earlier) of the physicality of sound-source. 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 addition, and not easily disregarded, the onset (or attack) properties of a sound give us some indication of the cause of that sound – a physical blow, a scraping sound, a movement, a vocal sound-source. The onset or attack of a sound is often 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 continued phase of the sound, we are articulating what the sound has revealed. It is possible to give the most utilitarian sounds an apparent vocal provenance by very carefully grooming a vocal onset onto a non-vocal continuation. The vocal "causality" in the onset can adorn to the ensuing sound in the most unlikely cases.
Similarly, some artists seek out exact rational properties, 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 exact operatic or cinematic idea or method, may be viewed as a musical ideal and its realization in performance an adequate or poor reflection of this ideal object. Reality is set in stone, or print, in the exact properties of architecture, the unchanging fluency 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 mental attempts to hold this flow in our grasp 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 these tests have to be revised or rewriten, the proportions of our architecture recomposed, the appropriateness of our methods reassessed in the light of experience.

For me, sound-composition, is seeking the means o’ the everchanging flow of sonic events, attempts to grapple with the very essence of human experience.

STABILITY & MOTION: REFERENCE FRAMES

In general, every property of a sound (pitch, loudness, spectral content, etc.) may be ad infinitum mobile, or in motion, e.g., cyclically varying pitch (vibrato) may be accelerating in cycle-duration, while drifting in pitch-depth. Cyclic variations 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 diachronous music system which requires 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 necessarily reproducible. We can, for example, reproduce the duration, duration and range of a rising portamento and in practice, we can differentiate a start-weighted from an end-weighted portamento. (See Diagram 4).

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 5) with any precision would be difficult (if not immediately impossible). 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 transmitted to other media. For the first time, we have sophisticated control over sonic motion in many dimensions (pitch, loudness, spectral content or format evolution, spectral harmonicity/discontinuity and other). And we 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 see it as an inevitable or an observable property in relation to a set of reference values. These provide reference points, or reference tools, in the continuum of sonic possibilities. Thus, for example, any particular language provides a phonetic reference-frame distinguishing those sounds and consequent types to be regarded as different (and hence capable of articulating different meanings) within the continuum of possibilities. These distinctions may be audibly ("1" and "2" 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 word diphthongs are in fact defined by the motion of their spectral centres (Goodman, see Appendix 9). 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 at the "one-of-value", to make very subtle distinctions of pitch motion and spectral motion. They are vital in our comprehension of the phonematic and tonematic 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. 111
in the sequential domain. Stream interaction in an elementary sense can be heard in Steve Reisch's phasing pieces and in the obligation section of Viola-3 (see Chapter 11) but in neither of these cases is the material in the separate streams undergoing real morphic change (as in Viola-3) the streams move in space relative to one another).

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

Imagine also a theme situation in which we see the quite separate evolution of two characters with entirely different, but internally consistent, goals and desires. Then external circumstances compel 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 store of development for any sound composer who can grasp the concepts and mathematical fundamentals of morphic flow processes.

PORTULIDE: TIME & SPACE

The form of living organisms is the result of continuous processes of flow (growth) - they are morphic forms expanded in space. Even the calculated remains of six centuries (e.g. the shell of the Nautilus) tell us about the process of continuous growth which formed them, and any spatial proportions we may observe are generated by these continuous processes. time-flows revealed in space.

What strikes us as formally coherent about a tree is not some count and static proportion of branching points and branch angles, but the way in which a flexible and environmentally sensitive expression of some branching principle is realized through a process of branching flow. The overall shape of a particular flow pattern (its result being the current state of the tree) limits 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 internally coherent and a musical score is not the musical experience itself, but a set of instruments to generate this audible substance, just as the genetic code is a set of instructions to grow an organism. The symmetries of the D.N.A. 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 mental experience of time-flow.

The way in which we divide experience into the spatially coordinate 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 poles in which we attempt to control the flow of experience and history.
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 transform performance in Western Art Music and the textual fixation of analysis and evaluation, it is difficult for music analytic thought to take on board morphic forms. Such a vast literature of musical analysis exists, grounded in sequential properties. By contrast, there has been no means to analyze, let alone describe and analyze, morphic forms.

The combined powers of the computer to record sound and to perform numerical analysis of the data to any required degree of precision (indefinitely if we ignore technical difficulties here). For those willing to deal with precise numerical representations of sonic reality, a whole new field of study opens up. To those fixed 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-parametron which moves slowly before slowly settling on the goal pitch, can be differentiated from a linear parametron 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 decomposing rising parametron which accelerates away from a pitch, a “leading-from” or dispersive ascension.
2. A rising parametron which slows as it approaches the final pitch and from which gradually emerges a widening vibrato (also making the percept louder) – a common genre in Western popular music – a “leading to”, or morphic anacrusis.
3. A sense of morphic instability 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 transformed to other parameters (loudness variation, change of cyclic loudness variation (tremolo rate), formant motion, harmonic shift, density fluctuations 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 widen cease to be perceived as a tone event and becomes a pitch-ridge structure. We have a simple example of morphic modulation, taking us from (perhaps) Hilliard field percepts to pitch-ridge field percepts.

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

We can see the same mixed classification in the Indian ragas system where a raga is often defined in terms of both a scale (of static pitch values) and various metric figures often involving sliding intervals (rhythm-types). In general, the case of pitch-reference-frames, values lying off the reference frame may only be used in ornamental or satisfactory features of the musical streams. They have a different status to that of pitches on the reference frame (cf. “blue notes” as near, certain kinds of baroque ornamentation etc.).

A reference frame gives a structure to a sonic space enabling us to say “I am here” and no elsewhere.

People with perfect pitch can use the pitch space in this absolute sense, but for most of us, we have only some general conception of high and low. We are, however, determine our position relative to a given pitch, using the notion of interval. If the octave is divided by a node or scale into an asymmetrical set of intervals, we can tell where we are from a set of sets without hearing the key note because the interval relationships between the note closest as within the set-of-intervals makes 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 B).

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 continuous in musical contexts and I will postpone further discussion of these until Chapter 9.

Traditional Western music practice is securely wedded to pitch reference frames. In fact on many instruments (keyboards, fluted violin) it is difficult to escape. However, in sonor 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 and away from, such a frame. Computer control permits the very precise exploitation of this area of new possibilities. It is particularly important to understand that we can base pitch, and even stable pitch, without having a stable reference frame and hence without having a Riemannian sense of pitch in the traditional sense. (See Diagram 9). We can even imagine establishing a reference frame for pitch–metrical types without having a Riemannian frame – this we already find in turn language in China and Africa.
**DIAGRAM 0**

Stable pitches defining a reference frame.

---

**DIAGRAM 1**

Conventional crescendo.

---

Stepped crescendo.

---

Stable pitches not defining a reference frame.
(every pitch is different)
mathematics and the continuum of our sonic experience (re-called "sound"") is thus 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
breach the necessary curve of experience and come up against the limits of human perception and
which itself has limits to the accuracy required for our calculations. The world of pure ratios in the
sense of testing systems, or temporal proportions, is seen for the idealization that it is. We come
time-to-time with the fact, the reality of human experience, and the realities of rational calculation.

Nevertheless, what begins by 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 "nuances" of speech
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 Maestoso School of Symphonic Severe 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 stunning
development. But traditional notation gives no means to add detail to this simple state-interpolation.

(More recent sonoristic developments include staccato crescendo: see Diagram 1.)

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

The subtle articulations of pitch, continuation and special dynamic properties can be observed in the
vocal music practice of many non-European music traditions, as well as within jazz and in popular
music. Musicians hand in hand, by ear, can carry these morphic form information from one generation of performers to another. It does not get filtered off into a separate domain of "performance practice" separated from stable-property notations and subsequently developed.

THE MANIPULATION OF SOUND: TIME-VARYING PROCESSING

Just as it has been at pains to stress that sound events have various time-varying properties, it is
important to understand that for compositional purposes we apply to sounds may also vary in time in
stable 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 medium dominated by the notion 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.

When this is not the case (e.g. quantized sequencer music on electronic synthesizers) we quickly lose
the musical quality.

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 validity. In a similar fashion an "effective value"
used as an all-or-nothing process on a wise or instrument, is simply that, an "effect" which may
occasionally be appropriate (to create a periodic general acoustic ensemble, for example) but will not
sustain our interest in a compositional sense 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
generally increasing (but subtly varying) vibrato, applied to a fairly static sound, gradually giving it a
vocal quality; a gradual spectral sweeping 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
stable use of time-varying parameters.

THE MANIPULATION OF SOUND: RELATIONSHIP TO THE SOURCE

The infinite multifacility of sound makes raises another significant musical issue, discussed by Allan
Savours at the International Computer Music Conference at IRCAM in 1985. As we can do anything to
a signal, we can do anything to listening whether the source sound and the goal sound of a computational
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 intermediates sounds. In source-based music there is a tradition
to claiming that transformation which can be explained and who's results can be seen in a source, by
definition defines a musical relationship. This musical textual idealism is out of the question in the
new musical domain.

An instrument which replaces every half-wave by a single pole equal in amplitude to the
amplitude of the half-wave, is a perfectly well-defined transformation, but reduces sound to a uniform
cracking noise. Nevertheless, for complex real-world sounds, such cracking 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 nearly perceptible similarity or refinement, regardless of the methodological rigor of the process which transfers one sound into another.

Another important aspect of Sennett's talk was to distinguish between source-focused transformations (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 uncertainty) 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-digital signal to produce delay-time related pitch-geing). 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 computational technique emerges, e.g. pitch-shifting in 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 on the source, however, retain the same infinite potential that the infinity of natural sound sources offers 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.
In this context, intonation 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 organization can provide structural reference points. We no longer have the traditional hierarchy: pitch, 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 motif is more apparent in the context of the whole piece than in the context of the excerpt here) into which a moment pitched event is inserted. As the sequence proceeds the units are given spectral pitch through successive filterbank filtering with partials of increasing Q, and fundamental frequency in the highest common factor (HCF) of those partials to the pitch domain. Once the 'strokes' enter, the rhythmic patterning is lost in a dense texture, but the pitch-focus remains in the consonant field of this texture. In the foreground the 'strokes' are linked by a new device, the falling portamento shape whose origin we hear when the specifically time-stretched voice pears through the texture.

In Sound example 13.3 we begin with a vocal texture involving its density, which begins to rise in intensity. As it does so the texture who's set to produce a noise band which is subsequently filtered filtered to produce a multi-pitched harmonic sound. The rising tension of the texture is first perceived as a secondary property (as articulation). But it soon becomes the binding structural element, as both of the ensuing sections, transitions, and sound events gain up and dominate the harmonic spectrum, and not simply a matter of directly perceiving a fundamental frequency in a spectrum. Such a frequency may not 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 (simplified) voice sound by placing the old harmonics on the left loudspeaker and the even harmonics on the right loudspeaker we hear a single voice sound between the two loudspeakers. This happens because of a phenomenon known as ear mixing. When sounds from two different sources enter our ears simultaneously we need some mechanism to disentangle the spatial location of the sound from those belonging to the other. One way in which the ear is able to process the data relies on the micro-invariances (jitter) in pitch, or tensions, and loudness which all naturally occurring sounds exhibit. The partials derived from one sound will all join in parallel with one another, while those from the other sound will jitter dissonantly but also in parallel with one another. This provides a strong clue for our brains to assign any particular partial to one or other of the source sounds.

In the loudspeaker experiment, however, we have removed this clue by averaging the spectrograms of the partials coming from the two loudspeakers. Hence the ear does not unravel the data into two apparent sources. The voice remains a single percept. (Sound example 2.1).

If, now, we gradually add a different vibration to each set of partials, the sound image will split. The ear is now able to group the 3 sets of data into two sound streams and assign two different source sounds to
what it does. The odd harmonics, say 300, 300, 700, 900, will continue to impart a fundamental of 100 cycles but will take on a clearer type quality (clarity produces only the odd harmonics in the spectrum) and move into one of the loudspeakers. The remaining harmonics, 400, 400, 800, 800, will be interpreted as having a fundamental at 200 (in 4th, 4th, 8th, 8th) and hence become "viva" or "viva" higher, will arrive from the other loudspeaker. Hence, with no change of spectral content, we have generated 2 pitch percept from a single pitch percept. (Sample example 2.2).

SPECTRAL AND HARMONIC PERCEPTIONS 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 perceived frequency. Once this exact relationship is disturbed, this spectral fusion is disturbed (see new 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. As "A" and "C" played on two flutes (or on two plains) are two distinct mixed sounds, each having its own integrated spectrum. Each spectrum is interpreted with itself because its internal microstructures run in parallel over all its own partials. But these microstructures are different in those in the other spectrum. Within a single spectrum, however, partials might nearly correspond to an A and a C (at equal numerical proportions, subject to the tempered scale) but will also have exactly parallel microstructures and hence base our auditory perception of a much lower fundamental (e.g., an A = 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 "Harmonious" (capitalized as shown) to apply to the traditional concepts of relationships 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 pitch in the sense of containing "harmonic" partials in a simple relationship to one another. This sense of "having a pitch" i.e., being able to assign a pitch to a specific class like E-flat or C# is quite a different concept from the (idealized concept of pitch described here. We will therefore refer to that traditional concept as Hertzian, an abbreviation for pitch-as-related-to--a-stationary-summated-harmony.

Perception of Hertzian pitch 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 Hertz 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 Hertzian field, then a set of pitches approximable to a Hertzian field and finally the "same" set locked onto that Hertzian field. (Sample example 2.3).

THE STRUCTURAL FUNCTION OF INTERPOLATION

It is instructive to examine the form of tamed composer phrases to illustrate this multi-dimensional approach. The sound examples here are taken from Tongue of Fire (1992-94). In Sound Example 13.1 we begin with a vocal sound whose pitch is spectrally time-stretched with a gentle swoosh at its end. The prior context is that of voice sounds. Voice sounds themselves are only recognizable as such through a complex interaction of properties (pitch-timbre, pitch glide speed and range, formant and formant glide rate, noise types, general rate and mini-regulatory of sequencing etc.).

After the first downward portamento we arrive at a motion based on variants of this time-stretched voice-tail. Here the stroboscopic sound falls or rises to viva or vibrato articulated in numerous ways. Each event in this segment again vocally, but the spectrally time-stretched tail extensions are linked through their time-evolving specific type, pitch and loudness are the principle articulating parameters.

This leads to a varied recapitulation of the falling portamento with tremolo idea, but now the lackness of the tremolo can be so deep that the original continuous sound breaks into a succession of wood-like ceils. This is a sonic modulation from voice to "wood" and is akin to a key change in the total system. However, in is only a passing modulation. We switch back again into a section of variants of the time-stretched tail of the vocal sound.

This section ends with a true sonic modulation, the clarinets coming events firmly establishing a new sonic area, less vocal, but "wood-like". It is interrupted by a stripping shift back towards the voice-tail, but, in fact this section enacted, event-enveloped version of the tail occurs without the vocal initiation, thus understating the sonic "key change" which takes place. As this event closes, the wood-like events rise in pitch and density, establishing a high granular texture. The voice has now been left behind. We arc in a new sonic "key".

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

The section ends with yet another surprising sound-modulation as the vocal must return but out of a set of other transformations) goes over into a sound like random vocal grit.

In my description of this event I have tried to emphasize the structural importance of sonority by comparing it to vocal structure. I am also stressing the structural function of interpolation as a continuum-domain analogues of total modulation. There are clearly important differences. The sets of keys in the vocal system form a finite, discrete and cycle set of possibilities (see On Sonic Art) and the vocal system is not of 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 spectral perception in Chapter 9).

26

101
create scales of perception between definable points without parameterizing 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, enabling relationships between sounds (definitively or through implicit process work) which are both perceptible and musically potent in some sense.

Given this multi-dimensional space, we can generate the notion of "modulation" (in the sonic sense).

We are already aware that recombining pitches can fundamentally change its affective character. Music transferred from what we think of as sounds, for example, or within instrumental practice, a change in "expression" in the production of the sound stream (e.g. boxing 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 tuned to structural properties of the sounds and the control of these structural properties. The distinction structural/expression arises from the arbitrary divide created by the limitations of notation. Features of sound we have, in the past, been able to cause with some exactitude (Hitchch, duration) are opened up to "rational" control (or at least mandatory exploitation) by composers and commentators. Those which remain (virtual, or vague in the notation are out of reach of this rationalizable control. They do, of course, remain under the intuitive control of the performer (see Chapter 1) but this type of control seems 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 rationalization) must come to terms with intuition. With precise sound-compositional control of the multi-dimensional space, we can move from what were (or seemed to be) all-or-nothing shifts in sound-space to a subtly articulated and possibly progressively site-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 musically articulated of a multi-dimensional sound space within our everyday experience. Consider the many effective 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 restraint (story, accompaniment, 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 music 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 articulate space awaiting musical exploration.

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 the single pitch defined. (Sample example 2.8).

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

Remaining now to portamenti and considering shifting portamenti, it is possible to isolate such moving pitches to Hitchch if the portamenti have special properties, that is, if they are cease-weighted and the means located in a fixed HARMONIC field, we will assign Hitchch to the events, regarding the portamenti as ornaments or articulations to those Hitchch. (Sample example 2.9).

Finally, end-focused portamenti (often found in popular music singing styles) where the portamenti seem clearly onto values in a HARMONIC field, will be perceived as instrumental ornaments to Hitchch. (Sample example 2.10). Portamenti with no such weighting, however, will not be perceived as Hitchch. (Sample example 2.11).

The same arguments apply to falling portamenti and even more powerfully, to completely moving portamenti. (Sample example 2.12). 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 as Hitchch, or a set of HITCHCHs 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 Hitchch 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 Hitchch!

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

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

PITCH-TRACKING

It seems odd to early in this book to tackle what is perhaps one of the most difficult problems in instrument design. While treatments 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 every one us competing 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 recognize pitch if we can find a cycle of the waveform (a wavelength) and then correlate that with similar codes immediately following it. The pitch is then simply one divided by the wavelength. This is pitch-tracking by auto-correlation (Appendix p.76).
We can also attempt pitch-tracking by partial analysis (Appendix G7). Hence in the frequency domain we would expect to have a sequence of (finite) 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 mapped onto distinct channels. We must then separate the true partials from the other information by looking for the peaks in the data.

Then, as a 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 band of accuracy, any set of partial frequencies will have a higher common factor; e.g., partials at 1003, 203, 307 and 513.5 have an HCF of 1. We must therefore reject already 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 context of the source sound is (see Appendix G7). In such relatively straightforward cases pitch-tracking can be very accurate, with perhaps occasional orientation problems (the pitch can be assigned to the wrong octave). However, in the general case (e.g., speech, or synthesized 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 data from the frequency domain to come up with our most definitive solution.

The pitch data might be stored in the (equivalent of a) breakpoints 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, in what range of change should we have changed? (See Diagram 1).

If we are working on an Hfitch reference frame the task is, of course, much simpler. If we do not confine ourselves to such frames, to be completely rigorous we could issue 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-size-threshold 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-frame 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-tone, either diatonically, 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 size (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 reprocessed sound. (See Diagram 2).

**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 themselves to be controlled by an external architecture of hitters and rhythm along traditional lines. In particular, MFR technology (1994) motivates this approach 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 see traditional approaches in musical form-building. On the contrary a broad knowledge of ideas about modality, form, rhythm, rhythmization, and distribution, control, large-scale musical forms in general, 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 suggest that a detailed examination of the stylistic reasons for the evolution and styles of structures of description 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 music practice. Similarly, a study of the World's languages reveals the great range of sound materials that enter into everyday human sonic interaction 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 precision of recent Western Art Music, and the fact that these precisions are constructed into the instrument technology (from the tempered keyboard, to the keyed flute, to the MIDI protocol) it is easy to see music as in a two-dimensional (pitch/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 ignoring the analytic hierarchization.

Sounds may be organized into multi-dimensional relations sets in terms of their Hfitch and Hfitch—field, or their pitch and pitch—scale, their vowel sets, harmony type, onset density, vibrato—acceleration—type, etc., with each of these trenched in distinct perceptible orderable parameters. Not all available parameters are perceptually separable in every case, e.g., vibrato acceleration—type may be inseparable when onset—density is very high. But in all definable situations we have many such parameters at our disposal.

We may also organize sounds in terms of their overall holistic qualities. This approach is appropriate both for large time—frame events where separate classes of properties may not be distinguishable and for sound sets created by progressive interpolation between quite different sounds (e.g., "Broad" versus "Rich" bass sounds Example 12.5). This latter sound interpolation illustrates the fact that we can
There are also some fixed techniques that can be used in special cases. Once the pitch of a sound is known, we can transpose it up or down using comb-filter transposition (Appendix p86). 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. If we start with a sound whose waveform are at 100, 200, 300, 400, 500, etc with a fundamental at 100Hz, we are left with a waveform at 200, 400, 600, 800, etc with the same fundamental at 100Hz,news which 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 distort the contour of the sound (its 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 pitch. The pitch of the instrument does not have to be altered to transcribe the most subtle or complex pitch flows and fluctuations without necessarily being able to assign specific pitch values at any point. For example, the reiterations of portamento and vibrato are treated in a particular established vocal or instrumental idiom, or in a naturally unscared bird or animal cry, could be transferred to an arbitrarily chosen nonsystemical, nonvocal, or even synthetic sound object. It would even be possible to transpose between such articulation styles without any stage having a quantifiable (measurable) or notable 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. Instructing an instrument, whether in an orchestra or another group of similar objects (clarinet, oboe, saxophone, trumpet, wooden bars etc.), or with variable access to the same objects (flute, ocarina, violin,取得, braun valve) to perform similar sounds with different pitches to be produced rapidly and easily.

There are two problems which 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 usually pitch (as Hitchcock did). More importantly the majority of sounds do not come so easily packaged either because the circumstances of their production cannot be reproduced (a breaking of sheet glass... every sheet will break differently, nonoptem how much care we take), or because the details of their own production cannot be precisely remembered (a spoken phrase can appear at different pitches by the same voice inside a narrow range but, assuming natural speech inflection, the first details of articulation cannot usually be precisely remembered).

In fact changing the pitch on an instrument always involves some spectral compensation, e.g. low and high pitch on the piano do have a very different spectral quality but we have come to regard these discrepancies as acceptable through the ability of instrument design (the piano strings resonant in the same sound box and there is a relatively smooth transition in quality from low to high string) and the familiarity of tradition. We are, in fact, changing the pitch of the original sound, but producing another sound, whose relationship to the original is acceptable.

There are a number of refinements to the interval perception process. We may interpolate amplitude data and frequency data separately (at different times with different intervals) 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 p26).

There are also perceptual problems in creating truly inaudible interpolation between two recognisable sound sources. If we do not recognize the gap time and we need enough time for both sound 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, similar noise), if the seamless transition is to be achieved without introducing artefacts which either suggest a threat and seem physically (socially) in the sound, or even reveal the mechanics of the process of composition itself.

A more difficult problem is created by categoric switching in perception, e.g. in the transformation voice->bees, we tend to perceive ~ "voice = a robot? = 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 to the maximum uncertainty for the transition to be accepted. In the voice-bees transition in Vox 3, a very high, complex, 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 undercover the sudden categorical 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 spectroscopically intermediate between source and goal. If we also allow progressive time->spatial (so that duration of sound source and goal sound can be matched through interpolation) and we place non aesthetic restrictions (such as the goal of 'translucency' in the examples above) on the intervening sound-styles, 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 in about causing convincing, small perceptual changes in the resulting sounds - not about the internal mathematical logic that produces them. To achieve an approximately linear scale of interpolation along a set of sounds may require a highly non-linear sequence of processing variation 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 inter\$plicable. In the eating scene: imposed reverberation can help to fuse a sound-percept by giving an impression of 'spatial integrity' (this sound was approximately produced at a single place in a given space). More profoundly, spatial streaming will tend to separate a fused sound, in a sequence or listener-station, if one set of elements moves left and the other right, we experience aural stream dissociation.

A movement from moose to tree stereo can, however, be used to enhance, or 'dramatize' the process of dynamic spectral interpolation. In Vox 5 most of the voice->other interpolations start with a voice in moose at front centre stage and interpolate to a stereo image (crown, bass) which itself often moves off over the listeners hearers. Spatialisation and spatial motion hence reinforce the dynamism of the transition.
Another way in which we can achieve interaction between the spectra of two or more sounds is via spectral masking. Here we construct a gest 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 these source has prominent high-frequency partials, it may mask out the high frequency deas in the other source(s) and the high-frequency characteristics of the masked sounds will be more clearly revealed if the first source passes, or gaps quieter. These aspects of the spectra of two (or more) sources may be played off against each other in a compositional interaction of the spectral data. This technique should, in general, be regarded as a form of neurally interactive mixing, rather than interpolation in the true sense. However, these source sounds have usually high partials which are strongly pitch-related, alterations in the spectrum of source sounds will produce interpolated spectral states. With voices, or other sources with variable form, this interaction will be particularly potent. (Sound example 2.12.)

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 Vol 5.

In Appendix p52 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 spacing.

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 sound (sound 1) to the second sound (sound 2).

It is important to understand that we are interpolating over the difference between the values in successive windows. We are moving gradually 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 (forward in window N+1) of sound 2. The latter process, lets us hear transitions between two static spectral states, would produce merely a spectral glide perceptually disynchronous from both source sounds. (Sound example 13.13.)

Our process, in contrast, is moving from the 'whobling' of one source into the 'whobling' of the other. For this very reason, our 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 naturally 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.)

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. This is a very complex task, and adopting this approach for every sound we can, would make sound composition unfeasible. Therefore 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 mixing 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 a 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 ergonomics sounds (speech, mouth-space) or moving sounds (e.g. portamento) this changes their perceived speed. (Sound example 2.9.)

Computer control makes such 'tape-speed' variation a much more useful tool as we can precisely control the speed change trajectory or specify a speed-change in terms of an initial velocity (tape acceleration). The availability of time–variable processing gives a new order of compositional control of time–varying processes. (Sound example 2.10.)

Waveret transposition is an unconventional approach which avoids the time–duration involved in tape-speed variation and can be used for integral multiples of the frequency. Here, each waveret (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 5 and Appendix p51.) This technique is very fast to compute but often introduces strange, signal dependent artefacts. (i.e. varying with the signal's details). It can therefore be used as a process of constructive distortion in its own right (Correcting the waverets in pairs, triples etc., before reproducing them, can affect the integrity of reproduction of the sound at the new pitch (see Diagram 6). The growing to choose again depends on the signal.

With accurate pitch-tracking this technique can be applied to true waverets (deduced from a knowledge of both the true wavelength and the zero-crossing information) and should avoid producing artefacts. (Diagram 9.)

The technique can also be used to tune the sound downward, replacing N waverets or waveret-space by just one of them, stretched, 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 waveret–domain approach is through brashage. To lower the pitch of the sound, we cut the sound into short segments, how, down in a tape-speed variation, which lengthens them, then splice them together again so they overlap neatly to retain the original duration. It is crucial to use segments in the prime-time frame (see Chapters 1 & 5), to use a segment is long enough to carry instantaneous pitch information, but not long enough to have a perceptible interval of space which would lead to unwanted echo effects within the pitch–changed sound. This technique, used in the harmonic, 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 reduced to 1/3, therefore frequency x 3, hence transpose up by interval of a 13th.

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.
If we have a very rapid sequence or a texture-stream we may achieve a dynamic interpolation between two quite distinct states by careful elemental 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 form a broad band noise granule stream onto a single pitch which we might intimate with formant-glides and vibrato articulations reminiscent of the human voice (formant synthesis: see Appendix). (See Example 12.10.)

**VOCODING AND SPECTRAL MASKING**

The distribution of partials in a spectrum (a spectral form) and the spectral contour (formants) are separable phenomena and impinge differentially on our perception. As discussed in Chapter 3 the spectral form creates our sense of the overall contour contributed by 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 sax), 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 "clip" on the second source, this process works best if the second source has a relatively simple 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 reconstructed by driving a generating signal through these time-varying filters. A sequence of constant sawtooth wave bursts of appropriate pitch (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 reconstruct 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 buzzes), or to change the vowel-sound characteristics (choice of buzz or noise). For interpolation applications, the signal we send through the filters will be one second source for interpolation (i.e., the one that's not the voice). It will usually not have a flute, or even a stable spectrum but can enhance the formant transfer process by "whitening" the spectrum, 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 recording procedures may, however, be attempted, using spectral data extracted by phase vocoder analysis).

This type of interpolation might also be applied progressively so that "voices", or lack of it, emerges out of a continuing non-vocal sound. (See 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-transition is applied (using a choral conductor!). If the spectral type of the sounds within the metered stream is made to vary appropriately (e.g. lack of syllables initially, 'T' and 'M' syllables at the wave peak, 'T' syllables 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 one "sea" construct - voices within voices.

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 sound duration. This is spectral shifting. (See example 12.13.)

**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 &quot;A&quot; will be found to be related to various peaks in the spectral contour. If we change the pitch at which &quot;A&quot; 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 12.8)

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 &quot;A&quot;-ness of the sound destroyed. (Appendix 12.6).

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

(a) Extract the spectral contour using linear predictive coding (LPC) (Appendix 12.13).
(b) Extract the partials with the phase vocoder (Appendix 12.11) or with a fine-grain LPC (see spectral-focusing below).  
(c) Filter the spectrum using the inverse of the spectral contour.  
(d) Change the spectrum.

(e) Reimpose the original spectral contour. (See Appendix 12.7).

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

Formant drift is an obvious problem when dealing with speech sounds, but none to be born in mind more generally. An instrument is characterized often by a single soundtrack (giving, vision) which provides a relatively fixed background spectral-tone 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 harmonizer 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 microvariations of the two pitches result fairly well in step, a situation impossible to achieve with two separate performers, though it may be closely approximated 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 H/Anomy to a melodic line. (Sample example 2:4).

As usual, the harmonizer algorithm introduces significant artifacts 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 see 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 I). (Sample example 2:15).

All these techniques can be applied dynamically so that the pitch of a sound gradually splits in two. (Sample example 2:46).

Small pitch-shifts, superimposed on the original sound add "body" to the sound, producing the well known "ch Ch" effect (as for "Chord") produced naturally by a chord of similar singers, or similar instruments playing the same pitches, where small variations in tuning between individual singers, or players, broaden the spectal band of the resultant sound. (Sample 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 only two pitches present in these cases but perceptions of conflicting frequency motions within the sound can certainly be established. With even careful tuning, including the phasing in and out of partials at the top and bottom of the spectrum, sounds can be created which get higher to E/Chiptech yet lower in texture (or lower in pitch but higher in intensity) - the so called Shepherd Tones. (See On Toxic Art and Appendix p7). (Sample 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 Tarus harmonica singing). (Sample example 2:9).

Another pitch phenomenon 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 in the relationship becomes 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 drift is upwards, even if the fundamental frequency is present in the spectrum and remains unchanged, the lowest perceived pitch may gradually move upwards. (Sample example 2:38).

The difficulties here might be compared with those in the process of in-betweening in animation. Here the director 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 must not be an skillful or decorative execution that results in the original frames but the end result must be intended to be the only consistent way. Similarly, we may impose the overtone (or wavecycle) shape of one sound on the overtone (or wavecycle) length of the other. With simpler sources, e.g. square wave to sine wave, this will have the pitch of the latter on the spectrum of the former. With complex sources, using waveforms, the result is once again complex and perceptually unpredictable. Again, we have natural transformation, but not true interpolation (waveform blending). (Sample example 2:39).

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 make stereo analysis windows obtained from the spectral analysis of two sounds.

When we produce a special analysis of a time-varying sound, we divide the sound into very time-slices, or windows, and analyse the spectrum in each window using e.g. the Fast Fourier transform. (see Appendix p2:2-5, 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 clearly welded "mix", if not exactly a factual mix. 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. (Sample example 2:34).

A similar concept can, however, be more easily achieved through transposing where we also have control over the segment size and are not therefore tied to regular oscillations. Two or more sounds can be transposed together (multiple-source transposition) with control of segment size, search range, pitch shift, spatialisation of segments etc, just as in single source transposing. This approach perhaps works best in producing a process-focused transformation (see Chapter 1) (e.g. granulated or pitch-spread) in which the two (or more) sources "pass through" the distorting grid of the process simultaneously. The process's artefact grille disguises the lack of true fusion of the sources - both are an integral part of the resulting "mix". (Sample example 12:7).

We can go one step further and attempt to integrate the wave-cycles, or waveforms, of the two sounds. Interfering wavecycles or waveforms will produce spectrally unpredictable sidebands; two interfered streams 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 modal modal distortion. This becomes a mutual transformation process rather than a true interpolation (interwave interference). (Sample example 12:8).

Similarly, we may impose the waveform (or wavecycle) shape of one sound on the waveform (or wavecycle) length of the other. With simpler sources, e.g. square wave to sine wave, this will have the pitch of the latter on the spectrum of the former. With complex sources, using waveforms, the results are once again complex and perceptually unpredictable. Again, we have natural transformation, but not true interpolation (waveform blending). (Sample example 12:9).
The third approach focuses upon the process of change itself. In Von Ses transformation sound→voice→scene aims to achieve clear recognition of both source and goal, and a seamless transition from one to the other without any intervening anacrusis which might suggest some other physical/visual, or even auditory, the technical process involved. (Sound example 12.2).

Moreover the dynamics of the change is a crucial parameter; in the voice→scene example, the voice almost melts slowly into the bee swarm; other transformations in Von Ses are quicker and more forceful, suggesting a generative space in the scene, getting or losing out the goal sounds – and this dynamics is enhanced by spatialization in the sense of spatial notice, or the emergence of voice images from a scene source. Here, the technical problems are those of simultaneity in the spectral transition and achieving the right dynamics, especially as interpenetration occurs often requires a relatively long time-frame in order to work smoothly. Musical context can play a vital role here.

INRETWEENING

The most obvious way to achieve some kind of interpenetration between two sounds would seem to be to mix them in appropriate (ative linings) proportions. However, as we know from our everyday experience the atone never creates perceptual fusion of the two sources. Our brain manages to unnotice the many sound rippling 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 sound synchronicity (for 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 another source.

Successful interpenetration by mixing can only be achieved if we can define either, or both, of them. In sounds with continuity, the very precise (to the sample) synchronizations of attack (attack synchronization) can achieve an instantaneous fusion of the sound image which is however immediately contradicted by the continuation of the sounds in question. (Sound example 12.4).

To achieve a completely convincing sound intermediate between the two, we must work with sounds whose continuations are (almost) identical, or with gate fringe-free sources (which have no continuations). In the latter case in particular, we can achieve good intermediate perceptions simply by mixing, but not necessarily. The ideal case (that the scene sound is transformed into the goal sound by a diversion process that returns the duration and general shape of the other sound to the level of the voice or voice (see distinctiveness criteria in Chapter 3). Superpositions of these two sounds in various proportions will create convincing intermediate states (interweaving) (Appendix p66). (Sound example 12.6).

Where the sounds have continuation, this process may be unsatisfactory. Consider, for example, the goal and source sounds of the sequence "Koo-ou" to "Blefl" 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 convincing set of intermediate sounds. The process of spectral overlap, successively optimized, has gradually separated the spectral components (and hence altered the spectral source) to a point where they will not simply fuse together again by mixing.

TETSTUTRA CHANGE OF UNPITCHED SOUNDS

The techniques of pitch change we have described can be applied to unpitched without any definite pitch. Inharmonic sound will, in general, be mapped in a way into pitched sounds. However, the source will usually change the mass→energy (the "pitched→tremolo") of noisy sounds so that they appear to move higher (or lower). (Sound example 12.13). The harmonics will usually have the same effect. Splitting the spectrum of a pure sound, noise-based sound using spectral shifting may not have any perceivable perceptual effect on the sound, even when the split is quite radical and the energy is often unintelligible as the spectrum is already full of energy. However, the problem of formant shifting when transposing will apply equally well to, e.g., registral speech sounds. (Sound example 2.22).

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

PITCH CREATION

It is possible to give pitch qualities to texture unpitched sounds. There are two approaches to this. One, at a higher level, is to add a small number of harmonious pitch to the source sound. It can be used for a filter not only to remove or attenuate parts of the spectrum but, by inversion, to accentuate any that remains.

Successful interpenetration by mixing can only be achieved if we can define either, or both, of them. In sounds with continuity, the very precise (to the sample) synchronizations of attack (attack synchronization) can achieve an instantaneous fusion of the sound image which is however immediately contradicted by the continuation of the sounds in question. (Sound example 12.4).

To achieve a completely convincing sound intermediate between the two, we must work with sounds whose continuations are (almost) identical, or with gate fringe-free sources (which have no continuations). In the latter case in particular, we can achieve good intermediate perceptions simply by mixing, but not necessarily. The ideal case (that the scene sound is transformed into the goal sound by a diversion process that returns the duration and general shape of the other sound to the level of the voice or voice (see distinctiveness criteria in Chapter 3). Superpositions of these two sounds in various proportions will create convincing intermediate states (interweaving) (Appendix p66). (Sound example 12.6).

The sounds have continuation, this process may be unsatisfactory. Consider, for example, the goal and source sounds of the sequence "Koo-ou" to "Blefl" 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 convincing set of intermediate sounds. The process of spectral overlap, successively optimized, has gradually separated the spectral components (and hence altered the spectral source) to a point where they will not simply fuse together again by mixing.

In this process works best on moving sources because here the energy will be distributed over the entire spectrum and wherever we place our filter (variable), we will be sure to find some signal components to work on. Very often Q on the filter will produce oscillographic pitch, while more intense Q and broader bands produce a vibrant focusing of the spectral energy. There are many degrees of "pitchness". "Pertinent" noise and a ringing oucitron (Sound example 2.34). We can also force our source through a stack of filters, called a "filter bank", producing "chords", and, increasing Q with time, move from one sound towards such chords. (Sound example 2.35).

In a sound with a simpler spectrum, a narrow, tight filter may simply miss any significant spectral anomalies — we may end up with silence.

A very sophisticated approach to this process is spectral focusing (Appendix p67). In this process, we first make an analysis of a sound with a spectral time-varying pitch using Linear Predictive Coding (Appendix p12-13). However, we vary the size of the analysis windows through time. With a normal size analysis window we will extract the spectral content (the formants) (see above). With a very fine analysis window, however, we will pick out the individual partials.
However, a percept of interpolation is possible without such clear differential classes 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 fixed metal sheet sound, and then imposing a
rapid series of hard-edged loudness increments on the resulting continuous, we move from the sense of
forced continuation of an elastic medium, to a sense of striking a hard inelastible medium. Both
physicality and causality have been altered. (Sound example 12.3).

Hence modifications to event 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.

MEDICATION : 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 such 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 Jungfrau where the ‘pure’ pitched singing
voice of a young boy and pure (pitched) sine tones are mediated through a set of intermediate pitched
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-recognizable) sounds, has a metaphysical undertone (a religious conception of ‘entry’ in the
context), 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 spans the range ‘boy’s voice’ to ‘sine tone’. These are sequentially
articulated according to an entirely different logic (a serialist sequencing aesthetics), which also governs
the rest of the musical organization in the piece.

A second approach to sound interpolation stresses the ambiguous implications of the sounds thus
created. Roger Reynolds 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 use of the interpolations, where we are most
uncomfortable about whether we hear in 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 interpolations without entirely losing
‘source credibility’ (i.e. is it anything at all that we can recognize?). This can be particularly difficult,
even with the most advanced technology.
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 spectra stretching, or time warping or filtering or randomization of perceptual elements etc) we will produce a sound which we recognize 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 source 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 spatially 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 white to a relatively short tone, pitch-scalable sound (e.g. white noise) so that it eventually involves such extreme and rapid change 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 concepts – a pure 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 some 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 recognizable in some referential sense. The source is a term, the goal is a visitor; 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 (bells) or it may seek out an ambiguous ground in which two recognition concepts conflict or cooperate within the same experience (the talking seal).

If we now use the analysis thus 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 finest characteristics of the original sound on the noise source (e.g. analyzed vocal speech will produce unvoiced speech) but where the window size was very flat, 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 flat, our output sound would vary from formant-shaped noise to strongly pitchshaped sound (e.g. focus 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 completely 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, whereas sound we input to the system. This technique is known as comb-filtering. Longer delays will give lower and less well defined pitchs. (See sound example 2.28.)

Both these techniques allow us to produce dual-pitch perception 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 portamento is an even simpler process. When a sound is reined with a very slightly time-stretched (orsheets) copy of itself, we will produce a gradually changing delay (See Diagram 6). If the sounds are start-synchronous, this will produce a downward portamento. We may work with more than two time-varied copies. (See sound example 2.27.)

Phasing or Phasing, often used to random noise, relies on such delay effects. In this case the signal is delayed by different amounts in different frequency regions using an all-pass filter (Appendix 9) and this shifted signal is allowed to interact with the unchanged source.

The production of pitch-motion across an appropriate place to end this Chapter as it stresses once again the difference between pitch and pitch and the power of the new compositional tools to provide control over pitch-in-motion.
CHAPTER 3

WHAT IS TIMBRE?

The special 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. piano, string, brass), however, in the more recent analogous studio, composers came into contact with oscillators producing frequency pitches, noise generators, producing frequencies noise bands, and "envolpe generators" which added simple loudness impatios 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-oscillation 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 fundamenta) and, provided this relationship persists for at least a grass-sline time-frame, the spectrum forms and we have a specific (possibly gliding) pitch. If the partials are not in this relationship, and provided the relationship (time to time) remains relatively stable, the ear's attempts to extract harmony (whole number) relationships amongst the partials will result in one hearing several pitches in the sound. These several pitches are not treated the same micro-oscillations and hence will be fused into a single percept (as in a bell sound). The case 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 "how" being gradually specally stretched: (Appendix y/19). This means that the partials are moved up the 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 indistinguishable "our" around it, akin to gliding, 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 harmon/crac/cromatic transition. As the process proceeds, the tail of the sound to 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 specimen is already bell-like. The syllable "hoow" begins with a very short broad bass spectrum with lots of high-frequency information ("V") corresponding to the initial ching of a bell. This leads immediately into a steady pitch, but the word formant is varied from "o" to "n", 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 bell 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 V2 3) illustrate the importance of the time-varying structure of the spectrum (not simply its loudness trajectory).

This suggests the ideal approach to time-stretching. In this case we may generate events at the original density, but for a longer time, then impose time-distorted field-variation parameters on the stream. Here the loudness trajectory, the pitch-range variation, the formant changes, the transitions to remain sound would move more slowly, but the overall density would remain as before. This is the temporal equivalent of granular time-stretching by granulation. (Sound example 11.38).

We here begin to touch upon interesting music-philosophical ground. For in the last case, the instance stream is clearly not a "pure" sound-event, in the same spirit we speak of this in Chapter 1. The instance stream is an example of a class of sounds with certain definable time-varying properties, just as a saw, 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 in its holistic characteristics. Composition focuses perception on what is being perceptually organized 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 system.

We have hence given these 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 instance stream, we may imagine another option in which the texture components 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 remains unchanged. This might better be regarded as a time-varying spectral-repetition-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 phenomenon within the sound percept and we may choose amongst the many compositional options available to us.
TIME-STRETCHING OF TEXTURE-STREAMS

We have left the discussion of texture-streams until the last because it introduces further
sub-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 perceivable time-varying field properties of the
texture-stream will thereby be time-stretched e.g. loudness trajectory, pitch-bend with change etc.
The resolution of the inner structures of grains may even alter the field percept (e.g. siren elements
becoming infrasonic sounds, or HF hitcher appearing as gliding pitchers) of the source sound. (Sampled
example 11.33).

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

Brassage techniques with above-granular segments, using regular segment size and zero
stitch-range (see Appendix 4ab) will quickly denature the unpatterned quality of the texture-stream,
as brassage repetitions introduce a "signature" 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 (e.g. directional change of the HF hitcher, loudness trajectory etc.) (Sampled
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 recombining the
original loudness trajectory in exactly the same time-land as is in the source texture-stream. We will
discuss this parameter separation further below. (Sampled 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 layer consistencies) 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-land of this sort, any time-varying
field properties (pitch-bend 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. (Sampled example
11.37).

DIAGRAM 1
We may vary this spectral stretching process by changing for 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 p.19). (Sound example 3.2).

Different types of stretching 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. No (as we have seen from the "no-no" example), does it mean that ---. Very short inharmonic sounds will sound percussive, like drums, annately coloured drums, or slurs to ---. (Sound example 3.4). These inharmonic sounds can be transposed and caused to move (suckle 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 quite fast, between the harmonic and 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 since those unaccented when pitch-shifting (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 p.17).

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 or a range 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 p.10).

As we know from singing, and as we can deduce from this diagram, the frequencies of the partials in the spectrum (determining pitch, harmonicity-inharmonicity, quality) 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 articulator time-varying control over spectral contour (none of the few examples in hand manipulable brass mute), the concept of formant control is lost familiar as a musical concept to traditional composers. However, we all use articulate formant control when speaking.
time—warping) allows us to produce many versions of such a phrase with different internal time proportions and perhaps, spectral emphasis, just as we might produce versions in the hip hop domain, of a rhythmic 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 completely spectrally evolving sound. Similarly, unadulterated continuation properties will cause slow glides with quite different musical implications to their underlying sources and we can alter these implications through time—warping. (Sample example 11.29.)

The time—warping of rhythmic vocal events provides us with an entirely new area for musical exploration. It is already possible to work with multiple fixed time—phases (e.g., mutually synchronized click—tracks in different tempos, as in Visa 2) or with time—phases mutually varying in a linear manner (as in the "phasing" pieces of Steve Reich, which generate new rhythmic, or melodic, periods through finely controlled delay) but we might also produce mutual interaction between rhythmic streams which are themselves trembling of deaccelerating, from time to time forcing the streams to pulse—synchronize in the same (possibly periodic) pulse—grouping unit (bars, cycles or note). (Sample example 11.30.)

This device is discussed in Visa 5) where these copies of an unaltered voice are slightly time—varied in different ways and made to more differently in space. They begin in synchronization, in a slight spatial location, and move to a similar stream at the end and by appropriate inversions of the speed variations (see Diagram 3). Linear speed divergence and rephasing becomes elements in the emergence and merging of such streams, or "countermerging", the time—fluid currents of counterstreams. Note, however, that the sound constituents of two time—varying counter streams do not have to be the same, or even similar. (Sample example 11.31.)

In fact, if the stream constituents are identical, are synchronized at some point and have slightly different (time—stretch parameters, their interaction produces periodic rising or falling from the point of synchronization (see Chapter 2 on Pitch Creation). (Sample example 11.32.) In the previous case (Sample example 11.33) these precise—delay pitch—units were avoided by carefully fading two of the three streams just before such artifacts would appear. The interaction of time—warping pulse—streams bears the same relation to final—tempo polyphony as controlling pitch—glide textures does to hip hop field pitch organization.

Time—stretching may lead to spectral fusion (see Chapter 3), e.g., Spectral time—stretching of noisy sounds by large amounts may produce sounds with gliding inharmonic spectra (Sample example 11.11) — or to a constructive dissonance (e.g., superimposed crooked piano) (Sample Examples 11.7 & 11.8) and spectral time—stretch reveals window—stepping in spectrally traced sounds (Sample Example 11.16). In both cases a long time—lengthened sound spot may not be perceptually reliable to the source and will require mediating sounds (e.g., less time—lengthen 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 greater—information (a comprehensible spoken sentence can become a rapid—fire, spectrally irregular sequence of glides) and continuation compressed into the indivisible qualitative percept of glue. Once again, if a perceived connection between source and glue is required (to build musical structure), mediating sound types may be necessary.

It is possible to extract the (time varying) spectral contour from one signal and impose it on another, a process originally developed in the analogues studies and known as warping (see connection with the phase vocoder). For this to work effectively, the sound to be warped must have energy distributed over the whole spectral range so that the spectral contour to be imposed has something to work on. Vowelling hence works well on noisy sounds (e.g., the sax) or on sounds which are artificially prewarped by adding broad band noise, or subjected to some noise producing distortion process. (Sample example 3.7.)

It is also possible to reposition the contour before imposing the new contour. This process is described in Chapter 2, and in the order preserving spectral manipulation in Appendix p.17.

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 focusing to frame certain aspects of the spectrum at a particular moment. We hold the frequencies of the partials, allowing their loudness to vary (speech). 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", as with, say when we might have anticipated that the frequencies would create this percept more directly. (Sample 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 glues, 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 ambivalent tans ("v", "d", "p", "k", "t", "t" the variety amongst cymbals and snaredrums give the lie to this.

Noisiness can be a matter of degree, particularly as the number of added components in an inharmonic spectrum increases gradually to the point of noise saturation. In consequence, any form of noise—wise in time: whispered speech is the ideal example. It can be more or less focused towards music or moving pitches, using broad—pass filters or delay (see Chapter 2), and it can have to own complex internal structure. In Sample example 3.9 we hear portamentos 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 noises", the type of cracking signal one gets from very poor radio reception tuned to no particular station, from motors of broad—band click—like sounds (either in regular layers — cacada — or irregular — masses of breaking twigs or pebbles falling onto this — or semi—regular — the gritty vocal sounds produced by water between the tongue and palate (e.g., Dutch "pht") or from extremely time—contracted speech stresses. There are also fluid noises produced by portamentos components, e.g., the sound of water flowing in a wide stream across many small rocks. These shake off into the area of "pterodactyl" which we will discuss in Chapter 8. (Sample example 3.10.)

These examples illustrate that the rather dull sounding word "noise" hides whole worlds of rich sonic material largely unexploited 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 trailing (see below). (Sampled example 3.11). Conventionally, all sounds can be assumed to sound with a noise-spectrum if superimposed randomly in a sufficiently frequency-dense and time-dense way. As the end of Sound-excerpt 3.9 the noise band finally recovers into the sound of voices. The noise band was in fact simply very dense multiplicity of very many voices.

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/time and time-fluctuating-spectrums/noise. (Sampled example 3.12).

A fruitful approach to this territory might be through spectral freezing, described in Chapter 2 (and Appendix II). This allows us to extract, from a pitched sound, either the spectral contour only, or the true partials, and so to 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 sound, to a more randomization of the partials of the original sound (and hence of the original sound itself). We can thus produce flexibly between these two states by varying the amplitude window size through time. This technique can be applied to any source, whether it be pitchily projected (harmonic), or inharmonic, and given us a means of passing from articulation noise to articulate noise-sounds in a seamless fashion.

Many of the sound phenomena we have discussed in this section are complex combinations of simpler units. It is therefore worthwhile to note that any arbitrary collection of sounds, especially mixed in mass, has a well-defined time-weathering specimen – a massaged group of echoes at a party; a sound of an orchestra, orchestras, individually, practicing their difficult passages before a concert. At each moment there is a composite spectrum for these events and any portion of it could be grit for the mill of sound composition.

SPECTRAL ENHANCEMENT

The already existing structure of a spectrum can be utilized to enhance the original sound. This is particularly important with respect to the onset portion of a sound and we will have discussion of this until Chapter 4. We may reinforce the total spectral structure, add additional partials by spectral filling the sound (without changing its duration) (Appendix II) and excising the shifted spectrum on the original. As the digital signal will retain its frequency purity, all the components in the shifted signal will line up precisely with the noise-sourced shifts and the spectrum will be thickened while retaining its (found) integrity. Octave enhancement is the most obvious approach but any interval of transposition (e.g. the intire) might be chosen that process might be expanded and the relative balance of the components adjusted as desired. (Appendix II). (Sampled example 3.13).

A further enrichement may be achieved by mixing an already dense spectrum with a pitch-shifted version which is left-right inverted. Theoretically this produces mostly a music-centre resultant specimen but in practice there appear to be frequency dependent effects which lead the resultant sound a new and richer spatial "fulminam". (Sampled example 3.14).
We may also extend grain-streams by duplicating elements, (granular time-stretching by grain duplication) e.g. imitate the original note at twice the initial separation and repeat each grain halfway into the new grains. This is a grain dimension analogue of wavelet 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 observed by a very narrow range pitch-scrambling 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 note-values, without thereby altering the (average) pulse-rate (density) of the stream. (Sample example 11.19).

TIME-SHRINKING

All the above processes may be applied to time-contractors, with noticeably different results. Type-speed variation time-contractor 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 irreversibly lost, i.e. if the grain is now time-dilated once more, the resulting sound will be considerably less time-vary significantly spectrally than the original. The process of successive continuations and expansion will give similar results to the process of spectral blurring (see Chapter 3). (Sample 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 duration approaches the lower grain time-frame boundary. (Sample example 11.21). If this process is repeated the impression of grains by reducing intergranular silence (where possible), or overlapping or splicing together the existing grains. These latter processes will tend to lose grains together. The process hence tends towards transduce-like loudness trajectory and eventually to a continuous percept, from which the original grain is not recoverable. (A new grain might be created by overlapping.) (Sample 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 waveform time-shrinking (or) duplication is replaced by waveform omission. This process also loses granular or sequential detail as the sound warps, though in a different way to grain separation contraction. We may choose to omit waveforms in a regular manner (e.g. every fourth waveform) in which case the sound becomes increasingly fuzzy, quieter, less detailed. Or, we may choose to omit the least significant (lowest amplitude) waveforms. In this case the sound retains its loudness but becomes less sequential detail. The first process results in a more prominent grain-stream trajectory, the latter eventually reduces any sound to a continuation-less but loud point sound. (Sample example 11.23).

THE CONSEQUENCES OF TIME-WARPING: TIME-FRAMES

As we have discovered time-warping is not a simple, simple process. The mechanical 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.

Finally, we can introduce a sense of multiple-sourcefulness (1) to a sound (e.g. make a single voice sound crowd-like) by adding small random time-changing perturbations to the loudness of the spectral components (spectral shuffling). This mimics part of the effect of several voices attempting to deliver the same information. (Sample example 3.15). We may also perturb the partial frequencies (Sample 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 close 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 relating to its morphology (time-varying shape) so that the resulting output will relate to the source sound through its articulation in time. (Sample example 3.17).

A filter may also be used to isolate some static or moving feature of a sound. In such ways, filters may be used to eliminate unwanted noise or hums recorded sounds, especially in 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 interfering relative loudness (spectral swiping) will eliminate the N most significant, i.e. quietest, channel components, with window by window. As an elementary level this can be used for signal-dependent noise reduction. But see also "Spectral Fuzzing" below. More radically, not of narrow band pass filters can be used to form a complex spectrum onto any desired pitch set (Harmonic field or the traditional sense). (Sample 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 on, or further compositional development. In particular, we can separate the spectrum into parts (passing band pass filters or spectral splitting) and apply processes to the N separate partials (e.g. pitch-shift, add vibrato) and then composit two the partials perhaps simulating the spectrum in a new forms. However, if the separate partials are changed too radically e.g. adding completely different vibrato to each partial, they will no sound when played, but we may be interested in the general disassociation 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 components, but also leaving unchanged the partials (partial routing). This is a much simpler task which will involve pitch tracking and pattern-matching the entire waveform where the partials might sit (on) a window by window basis. Ideally it is meant as one way in with information would (when the form of the spectrum is to be 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 reassembled sound and thereby provides a bridge between unique recorded sound-events and the control available through synthesis.

91
SPECTRAL FISSION & CONSTRUCTIVE DISTORTION

We have mentioned several times the idea of spectral fission where the parallel micro-oscillation of the many components of a spectrum causes us to perceive it as a stilled 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 ars of partials to be perceived independently — the single sound screen will split into two. (Sample example 3.19).

Spectral fission can be achieved in a number of quite different ways in the frequency domain. Spectral

\[ \text{distortion} \]

is a process that draws our attention to the individual spectral components by isolating, or emphasizing, each area of sound. This can be achieved purely visually by using appropriate vowel forms or freebys. The computer can apply this process to any sound-sound, even its own motion. (Sample example 3.28).

Spectral tracing strips away the spectral components in order of increasing loudness (Appendix a, p.28)

When only a few components are left, any sound is reduced to a delicate arista of defiled sine-wave constituents. Completely varying sources produce the most fascinating results as structures which are at any moment in the permitted group (the loudness) change from window to window. We hear new partials entering (while others leave) producing "inflected" intervalar to the sound source. This feature can often be enhanced by time-varying so that the rate of partial change is slowed down. Spectral tracing can also be done on a time-variable manner so that sound gradually dissolves into its internal sine-wave structure. (Sample example 3.21).

Spectral time-varying, which we will deal with 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 harmonic 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 insinuous window values. In general, these are insinuous and hence we produce a "metallic" insinuous (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 affected in a time-varying manner so that the insinuosity emerges gradually from within the stretching sound. (Sample example 3.22).

Alternatively we may elaborate the spectrum in the time-domain by a process of constructive distortion. By stretching for waveforms (zero-crossing points: Appendix a, p.28) and then repeating the waveforms before proceeding to the next (without wave-going) we may time stretch the source without altering its pitch (see elsewhere for the limitations on this process). (Appendix p.25).

Waveform correspond to waveforms in much pitched sounds, but not always (Appendix a, p.28). Their advantage in the context of constructive distortion is that very noisy sounds, having no pitch, have no true waveforms — but we can still segment them into waveforms (Appendix b, p.28).

In a very simple sound source (e.g. a steady waveform, from any oscillator) waveform time-varying produces no amelioration. A completely evoking signal (especially a noisy one) such waveform will be different, other radically different, to the previous one, but we will not perceptually register the context of that waveform in its own right (see the discussion of time-fractures in Chapter 1). It merely contributes to the more general percept of sonorities. The more we repeat each waveform however, the closer it comes to the sound threshold where we can hear out the implied pitch and the spectral quality implicit in

alter the sound percept by making the indivisible qualitative character of the onset become a time-varying percept (see Chapter 1). (Sample 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 can be achieved by time-variable spectral time-stretching (spectral time-varying), making the time-stretch equal to 1.0 (i.e. no stretch) over the first few milliseconds of the source and then increasing as rapidly as we wish to any large value we desire. (Sample 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 enclosure) 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 residuum. Clearly also, there are many combinations of onset transformation and continuation transformation. (Sample example 11.15).

In a very long spectral time-stretch of a sound's continuation, where the sound is specially varying, we can reveal these changes by stripping away partials from the time-stretched sound until only the few prominent (say) twenty remain (spectral tracing: see Appendix a, p.28 and Chapter 3). As the spectrum vanishes in few partials will cease and leave this honed sound and we will hear out the few entries as 'revealed moments', a type of constructive distortion. If the time-stretch is itself not time-varying, these revealed moments will be relatively pulses at the time-separation of the original window duration multiplied by the time-stretch factor. Spectral time-varying will create a fluid (gradual) tempo of entries and will accelerate into the time continuum where the time-stretch factor reduces towards 1.0 (no stretch). (Sample example 11.16).

GRANULAR TIME-STRETCHING

The time-stretching of grain-streams is problematic. We have seen, as we can time-stretch the sound of a sound we risk completely altering its perceived character. We overcome this problem by time-variable time-stretching (time-varying) a sound in such a way that the onset was not stretched. However, grain-streams act as 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-following to uncover the linearity 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, creating the sequence of grains to assistence, rhythmics, randomness scatter etc. (granular time-varying by grain separation). (Sample example 11.17).

With even moderately large granular time-varying of this sort, the sound character of a grain-stream breaks down in our perception — we hear only isolated lacunae events, the elements of potential musical phrases. Conversely, time-stretching 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 'point" in their own right. (Sample example 11.18).
Except for x2 time-stretching, when working on musically interesting sound sources, waveform repetition is more useful as a spectral sound transformation procedure. Successive applications of such x2 stretching does not significantly reduce the beat-mixing effect in long-time dilation.

SPECTRAL TIME-STRETCHING

A summed sound with stable spectrum and no distinctive onset characteristics may be analyzed to produce a (windowed) frequency-domain representation. We can thus recognize the sound, using each window to generate a longer duration (place vectors). The result 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 x4 time stretching. (Sample 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 that the human ear perceives the significant change from window to window (in fact a "step") as a smooth continuous transition. Once we do too long a synthesis from individual windows, e.g. a x4 time-stretch, the spectrum of each window is maintained long enough for us to become aware of the jumps. The continuity of the original source is not recommended.

One way around this limitation is to stretch the source x4, synthesize the result, reenvelope and time-stretch by x4 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-averaging). The procedure ensures a perceptually continuous result even at x4 time stretch. (Sample 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 down for us to hear the original source involved. In general we will have a resulting sound with a "glooming" (inharmonic/"metallic") spectrum in place of our source noise, at very great time-stretching factors. (Sample example 11.11).

Using time-variants 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 artifact in the sound source. Zooming in to a different point in the source will produce a different inharmonic artifact in the new sound source. 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). (Sample example 11.12).

The ultimate extension of this process is spectral foreshortening, where the frequencies or loudness 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 (true time-frame) onset characteristics, e.g. an attack-dispersed sound (piano, bell, piccolo) with an interpolated spectral time-foreshortening over a long stretch, will radically its waveform. With a 5 or 6 fold repetition therefore, the source sound begins to reveal a lighting fast attack transition, of all a slightly different spectral quality. A 52 fold repetition produces a clear "koomen nada or" apparently quite different from the source. A three or three fold repetition produces a "plashing" like aura around the sound in which a glimmer of the beam source is beginning to be apparent. (Sample 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 result is entirely different. The new sounds are time-domain amorphous consistent with the original signal, rather than realizations of an intrinsic internal structure. For this reason I refer to these processes as compositional defacement.

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 amongst existing 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 silence between a moving signal and a synthetic drone. (Sample 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 reconstrucing the signal. This can be used to "wash out" the detail in a segmented signal and works especially effectively on spiky crackly signals (those with loud, bright peaks). We can do this and also reduce the number of periodic spectral (time, & blue) 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. (Sample example 3.25).

Finally, we may shuffle the time-window data in any way we choose (spectral shuffling). Shuffling windows in groups of 4, 8, 17, 64, etc. With large numbers of windows in a shuffled group we produce an inedible rearrangements of signal segments, but with only a few windows we create another process of sound blurring, akin to bramaging, and particularly apparent in rapid sequences. (Sample example 3.26).

SPECTRAL MANIPULATION IN THE TIME-DOMAIN

A whole series of spectral manipulations can be effected in the time-domain by operating on waveforms defined as pairs of time-stretches. Basing in mind that there are not necessarily correspondents in true waveforms, even in relatively simple signals, we will anticipate producing various unexpected artifacts 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. At this process shapes the original form of the wave. I will refer to it as destructive distortion. The following manipulations suggest themselves.

We may replace waves with a waveform of a different shape but the same amplitude (wavelet synthesis: Appendix p52). Thus we may convert all the waves to square waves, triangular waves, sine waves, or even non-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 "shut" 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 wave-like pitch has waveforms replaced by square waves, and once by sine waves. Two interpolating sequences (see Chapter 12) between the "wave" and each of the transformed sounds is then created by interpolation (see Appendix p46 & Chapter 12). (Sound example 3.27).

Inverting the half-wave-cycles (wavelet inversion: Appendix p53) unusually 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 (wavelet distortion: appendix p52) (Sound example 3.28).

We may average the waveform shape over N wavelets (average averaging). Although this process sounds similar to the process of spectral blurring, it is in fact quite irrational, averaging the waveform length and the wave shape (and hence the resulting spectral content) in perceptually unpredictable way. More interesting (though apparently less promising) we may replace N in every M wavelets by silence (wavelet excision: Appendix p51). For example, every alternate waveform may be replaced by silence. Superficially, this would appear to be an unappealing approach but we are in fact thus changing the waveform. Again, this process introduces a slightly rising "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 waveform in any desired proportions (wavelet harmonic distortion) by making copies of the waveform which are 1/2 as short (1/2 as short etc) and superimposing 2 (3) of these or the original waveforms in any specified amplitude weighting. With an elementary waveform then the added harmonics are a natural 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 wavelets in any specified way (wavelet shuffling: Appendix p51) or reverse wavelets or groups of N wavelets (wavelet reversal: Appendix p51). Again, where N is large we produce a fairly predictable rearrangement of wave segments, but with smaller values of N the signal is altered in subtle ways. Values of N at the threshold of gram perceptibility are especially intoxicating. Finally, we may introduce small, random changes to the waveform lengths in the signal (wavelet shaking: Appendix p51). This has the effect of adding "noisiness" to clearly pitched sounds.

Such sessions procedures work particularly well with short sounds having distinctive loudness trajectories. In the two example a set of such sounds, suggesting a bouncing object, is destructively disturbed 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).
Braunage 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 generalized recombination and generative generation. (These possibilities are discussed in more detail in Chapter 5 in the section "Contracted Combination" and in Appendices pp44-45 and Appendix p53.) We need add only that using Generative Reorganization 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 more time-stretched and ends as a texture within the development of the source.

A more time-stretching satisfactory application of the Braunage process to sequences such as speech streams can be achieved by source-segment synchronous Braunage. 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 braunage-time-stretched and reassembled together in one operation. Because none of the goal segments crossover between source segments, we avoid artifacts created at source segment boundaries in simple braunage (see Diagram 1).

**Wavelet TIME-STRETCHING**

Time-stretching can be achieved by searching for zero-crossing pairs and repeating the waveform thus formed (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 wavecycle is not necessarily a wavecycle. As a result this technique will have some unpredictable, though interesting, sonic consequences. Sometimes parts of the signal will pitch-shift (e.g. at x1 time-stretch, by an octave downwards, as in tape-speed variation). For a x2 (or even x3) time-stretch, artifacts can often be reduced by repeating pairs (or larger groups) of waveforms (a special case of pitch synchronous braunage). These kinds of artifacts can of course be avoided, if truly pitched material, by using a pitch-following instrument to help us to distinguish the true wavecycles. (See example 11.6).

As the number of repetitions increases, other artifacts begin to appear. At x3 there is often a "plugging"-like oscillation of the sound. With time-evolving and noisy signals, at x6 time-stretch a rapid stream of pitched beats is produced as each wavepoint 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 fusion is heard "additionally" within the source even at x4 time-stretch. (Appendix p55). (See example 11.7).

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

**CONCLUSION**

In a sense, almost any manipulation of a signal will alter its spectrum. Even editing (most obviously in very short time-frames in braunage 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 "spectra". Spectral thinking is integral to all sound composition and should be borne in mind as we proceed to explore other aspects of this world.
also much more robust than the old analogue recording tape-speed variations, 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 variations 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, using 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 portamento. (Sound example 11.3).

BRASSAGE TECHNIQUES

As discussed previously, Brassage involves casting a sound into successive, and possibly overlapping, segments and then rearranging these by resynthesizing them together (i.e. past time-stretching, in exactly the same order) but differently spaced in time (see Appendix P44–AB). It can also be arranged, by appropriate choice of segments 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 harmonizers) 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 artifacts – 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-stretches the perceptual connection between source sound and goal sound may be remote and may require the perception of melodies (with less time-stretching) to make the connection apparent. Repeated application of Brassage techniques to a source (an 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).
CHAPTER 11

TIME-STRETCHING & TIME-FRAMES

Time-stretching and time-shifting warrant a separate chapter in this book because they are procedures which may breach 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: temporal, wavecyclic, status, and perceptual evolution". Furthermore, the degree of time-stretching of a sound may itself vary with 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 reach we desire. Below we will look at tape-speed variation, beat-squash techniques, waveform repetition (waveform time-stretching), frequency domain time-stretching (spectral time-reshaping), and grain separation or grain displacement (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 during the result at the standard sampling rate: see Appendix A). This approach is used in sampling keyboards (1983) and tape-playing instruments e.g. in Clavia.

In both cases the procedure (tape-speed variation) not only changes the sound duration but also the pitch because it alters the wavelength and therefore the frequency (see Appendix A) in the time-domain signal. Similarly, it changes the frequency of the partials and hence also shifts the spectral contour, and hence the timbre. And it time-stretches the event characteristics, probably radically changing the sound perceived in another way (Appendix B). (Signed example 1.)

Although time-warping, pitch-warping and formant-warping are thus not independent, this approach has its musical applications. In particular (real-world) octave (ms) upward transpositions can be used in short time-frame reiterations of a sound's overall 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 complimentary domains. Digital recording and transposition is

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 automatically tied to the way in which the sound changes. In this Chapter and the next we will look at these properties. In fact, the next chapter, entitled "Cessation" might seem to deal happily with all these properties. Why should we single out the properties of the onset of a sound, its attack, as being different to those that follow?

The onset of a sound, however, has two particular properties which are perceptually interesting. Most naturally occurring sounds the onset gives us some clue as to the casualty 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 help better 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 for 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. Then 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 sending some energy threshold, to push the system into motion. Bells need to be struck, drums blown etc. Some resonating systems can, with practice, be put into resonance with almost no discontinuity (the voice, bowl gongs, chimes). Others either require a transient onset event, or can be initialized with such an event, (drum or brass tonging). 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 (pano-swinging, bell). Other systems produce terminally short sounds as they have no internal resonance. Such sources can produce either individual short sounds (drums, xylophones, many vocal controllers) or be activated iteratively (drum roll, ‘re’, low confluence notes).

Onset sounds are a special case in which perceptual considerations enter into our judgments. Low and high confluence notes are both produced by the discontinuous movement of the rod. However, in the lower notes we hear out these individual motions as they individually fall within the given time-frame (see Chapter 1). Above a certain speed, the individual rod movements fall below the given time-frame boundary and the units meld to perception into a continuous event. Sounds which are perceptually complex, 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 combine with these. 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 change, a breathy release, or whatever, with the dimension 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 recognizability 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 sound recognition. Sound sources with internal resonance and natural decay (string piano strings, snare beats) are also partly recognizable 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 (kymograph lines, 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, resonances 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 la (on Diagram) will appear "sudden" as the trajectory implies that all of the energy is imparted in an initial shock and then dies away slowly. We can create the percept "struck object" by imposing such a brief loudness trajectory on almost any spectral structure. For example, a tone-pitched 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 "struck" sound may appear like struck wood (Sound example 4.1).

If these grains are then sparsely attired (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 "hard-like" or "cymbal-like" but we are missing 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 overhanging type 1b) which has a quasi onset and peaks in 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 coming on to 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 fasted scraping together of materials whose the evolution of the process is controlled by forces external to the natural vibrating properties of the material e.g. sawing or bowing, particularly when these produce forced vibrations rather than natural resonant frequencies. (Sound example 4.4).
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 hollowed 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 sounds with lots of high frequency components (hisses and other ringing metal sounds) where the approach of the microphone to the vibrating object picks up more and more high frequency detail. This zoom in 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 centimeters) in contrast to the typical range of the visual screen. (Sound examples 18.11).

PHYSICALITY AND CAUSALITY

As we have already suggested, loudness trajectory plays an important part in our attributions of both the physicality of a source (rigid, soft, loose aggregate etc.) and the physicality of the excitation (chuck, struck etc). The loudness trajectory of the sound onset is particularly significant in this respect, and often simple onset-exponential manipulation is sufficient to radically alter the perceived physicality of the source. Thus almost any sound can be given a smack-like attack by providing a sudden onset and then an exponential decay. Conversely, a strongly percussive sound may be softened by very carefully 'drawing off' the attack to give a slightly more slowly rising trajectory. (Sound examples 18.12).

This area is discussed in more detail in Chapter 4.
ALTERNED PHYSICALITY

When we proceed now to sounds which also have a continuation, subtle alterations of the sound characteristics may still radically alter the perceived physicality of the sound. For example we can impose a nodding—trembling (snack-like) onset on any sound purely by providing an appropriate constant loudness trajectory, or we can make a sound grating—buzzing by giving it an onset loudness trajectory which rises more slowly. In the grating time-frame we can provide from the "streak facilitation object" arising "stress moving object" to the "rabbit" and beyond due to that situation where the sound appears to run out of sound like the "singing" of bowl gongs or lutes. (Sound example 4.5b).

When the sound has a "strik" quality, we may imply not just the energy but the physical quality of the string medias. Harder stringing agents lead to higher frequencies in the spectrum of the vibrated material (compare padded stick, rubber headed sticks and wooden sticks on a "shripfrote"). We can generalise this notion of physical "harshness" of the 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 metaphorical terms. Grate time-frame events have an indivisible qualitative unity as perecepts. We can give physical and mathematical correlates for many of the perceived qualities, but in the immediate moment of perception we do not apprehend these physical and mathematical correlates. These are things we learn in appreciation 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 transpositions. The rate of decay is clearly a matter of aesthetic intent and judgement). (source stacking). The transpositions might be in the same octave of the original sound, or tone-combined (as with tape-speed variation; see Chapter 10). 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 ramps, 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 aural 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-representation spectrum suggesting physical softness and elasticity. We are not constrained by the photographically real, but our perception is guided by physical sensation even when listening to sound made in the entirely contrived space of sound production. (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 noise to "resolve" onto an actual wavelength of the modified sound by providing that wavelength with repetitions of itself which are increasingly randomised i.e. noise. This produces a plucked-string-like attack (sound plucking) and relates to a well known synthesiser instrument for producing plucked string sounds called the Kaymule Simplex. (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, rise—extending the onset of a sound will alter in loudness trajectory and may even extend it beyond the grain time-frame. As the onset is so perceptually significant, it can be used to control the perceived physicality of the sound which contrasts our experience of natural acoustic environments. Thus presumably recorded loud sounds may be played back very quietly, while distinctly quiet sounds may be projected with very loudness. These features of the sound landscape are discussed more fully in the Sonic Art.

In sound composition (in general, the loudness balance between diverse sources is described in a (possibly graphic) mixing scene, 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 grains size events. Balance changes within the grain time—frame is in fact a means of generating new sound substances intermediate between the constituents (e.g. shivering, see Chapter 12). (Sound example 10.8).

Mining 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 ofducking is used to this end. Here the general level of some of the other instruments is lowered in time 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 try 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 interweave a sound of more constant energy in large crowd talking (amongst itself, heavy trafic) 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 maximums are set and on the intrinsic impact 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 (each 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 enclosures 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 (in cold clear nights, sound waves are refracted downwards and hence sounds travel further).
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 it 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 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 brighten 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 (Sample 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. Each sound is made more strongly reverberant and louder events dominate. Combining this with triggered spatial position control we might utilize a spatially-stereo-versus-event-deepness 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 it is reduced, producing a more subtle containment of the sound. This process may be used more generally, setting the threshold level quite low, in flatter or smooth the general trajectory of a sound, thereby e.g. "calming" a hyperactive sequence of events.

An example, 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 general 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/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 coarsening 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 proportionality characteristics (close-miked, or miked at a distance such that room ambience is significantly incorporated into the sound).

time-stretching the course is much more perceptually potent than time-stretching the continuations.

This issue is discussed in Chapter 1. Also editing procedures on sequences (motilities, speech-stream etc) in many circumstances need to preserve event onset if they are not to radically alter the perceived nature of the material (the latter, of course, may be desired. Finally, extremely dense textures in mono will eventually stress sound characteristics, whereas stereo separation will allow the one discrete event stands out in very dense situations. (Sample 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 "vocalness" when continuation information contradicts our initial intuitive assumption. The piece Voices uses this causality transfer throughout in a very conscious way but it can operate on a more immediate level.

Linen 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 silence (i.e. a silence at a zero crossing: zero-cutting) in this sound where the vocal initiation is replaced into its non-vocal (but voice derived) continuation. On Voices the vocalization/vocalization transitions are achieved by smooth spectral interpolation, rather than abrupt offsetting. 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 vocal onset events, we may not so easily lose in the discontinuity.

A more radical causality shift can be produced by onset fissions. When we hear two sounds at the same time certain properties of the event allow us to differentiate them. Even when we hear two voices playing in unison, we are aware that we are hearing two voices 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 another but generally out of step with both of the other source. So in the combination we can usually separate the sources. Secondly, the onset of the two events will be slightly out of synchronism no matter how accurately they are played. Thus we can usually separate the two sources in the onset moments.

If we do this precisely the onset of two (or more) sounds to the nearest sample (event synchronously) our ability to separate the sources at onset is restored. The instantaneous percept is one of a single source. However, the continuation immediately reveals that we are mistaken. We then produce a percept with "false causality." At one output it is one source but it rapidly bifurcates into two.

In Sample example 4.11 from Voices the process is applied to three voice sources. Linen carefully to the first sound in the sequence. The percept is of "left" 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 sounds are also gradually spectroscopically stretched (See Chapter 3) becoming more and more bell-like in the process.
FINE-GRAINED ENVELOPE FOLLOWING

The result of the process of envelope following will depend to a great extent on how the loudness trajectory is analyzed. To assess the loudness of a sound in 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 p29). 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 rhythmic 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 modulations of loudness within a single syllable (i.e., 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 loudest level at the smallest meaningful time-frame (around 0.025 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 monotonous silence between the grains (corruption). This both allows the quality of the sound and makes the grains more amenable to independent manipulation (Sound example 8.5).

Following or manipulating the loudness trajectory below the grain time-frame has little meaning. The instantaneous value (instantaneous 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. Impressing loudness changes over one (or only a few) wavecycles, or wavelets, is effect to change the local shape of the wavelet (and is hence a process of spectral manipulation rather than of loudness control (see Appendix p29)). But if we gradually expand the number of wavecycles or wavelets we cause to fall under our loudness-changing envelope, we pass from the spectral domain to the time-domain of grain-structure and eventually to that of independent events. These transitions are nicely illustrated by wavelet enveloping. In Sound example 16.6, the number of wavelets 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 comprehension of sonic events.
ENERGY TRAJECTORIES

Sounds in the real world which do not simply the way to something require some continuous energy input such as bowing, blowing, strumming or an electrical power supply to maintain them. Usually the level of the flow of energy into the system determines the loudness of the sound 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, segmental or temporal sound by simply imposing a loudness trajectory, or envelope, on it (envelope). (Sound example 10.1.)

Such a trajectory may be derived directly from the activity of another sound (generating a piece of music) or a live performer (envelope following). Envelope following is particularly useful for creating a loudness accenture (a consummating onto a key sonic event), a particular musical device as most musical sounds have precisely the opposite loudness evolution. (Sound example 10.2.)

Conventionally, we may begin with a sustained sound (continuous, grain-stream, sequence, texture) and give it an exponentially decaying loudness, through enveloping. This suggests the sound originated from some vibrational medium which has been struck (or otherwise set in motion) and then left to resound, especially if the onset of the sound is reinforced. With simple sustained sound this can completely shift the apparent physicality and causality of the source (see below). With sounds which more easily IMPLY what goes (e.g. speech) we produce an interesting dual perception. (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 it (on the trajectory envelope transformation). As already suggested we might alter the trajectory to an entirely different sound (enveloping or envelope substitution). Sequences of events which are currently strongly characterized by loudness articulation may thus be "reinterpreted" with an entirely different sonic substance (Sound example 10.4.).

We may also modify the loudness trajectory we have extracted (envelope transformation) and supply it to the original sound. Note that we must do this by envelope substitution and not simply by enveloping (see Appendix 49h & 41). Thus we may exaggerate the loudness trajectory to heighten its energetic (or dynamic) evolution (expanding: see below).

The loudness trajectory may also be lengthened (envelope extension) or shortened (envelope contraction), extending or contracting the associated gestures (see for example the discussion of tone-prolonging techniques 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 progressively contoured (see below).

WHAT IS CONTINUATION?

Apart from grain-disruption sounds, once a sound has been initiated it must continue to evolve in some way. Only conceived 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 alike.

Some types of sound-continuation are, however, quite special. Sounds made from a sequence of perceived rapid events (grain-streams), sounds made from sequences of short and different elements (semaphore) and sound which dynamically transform one set of characteristics into a quite different set (dynamic interpretation) all have special continuation properties which we will discuss in later chapters. Here we will deal with the way in which certain simple 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 for grain-streams, sequences and dynamically interpreting sounds. They are not mutually exclusive.

DISPERSE CONTINUATION & ITS DEVELOPMENT

Certain natural sounds are initiated by a single or brief cause (riving, short blow or note) and then continue to evolve because the physical material involved has some internal momentum (unpitched material string sach, the sound is cushioned by a cavity resonator (one drum, resonant ball acoustics)). As the medium is no longer being excited, however, the sound will usually gradually become quieter (but inevitably; for example the sound of the lute-tman 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 continuations sound dispersal. (Sound example 5.1.)

In the studio we can immediately reverse this train of events (sound reversed), causing the sound to grow from nowhere, gradually accumulating all of its spectral characteristics and ending abruptly at a point of (usually) maximum loudness. The only real-world comparable experience might be that of a car's siren 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 starting if they are made from non-linear dispersions. The decay of loudness and spectral energy of a piano note is too close to linear, so the associated accumulation is little more than a crescendo. Gong, or tam-tams or other complex-spectra events (e.g. the resonance of the undamped piano frame where struck by a heavy wood object) have much more complex dispersal in which many of the initial high frequency components die away equally. The associated accumulation therefore begins very gradually but accelerates in spectral "increment" towards the end, generating a growing sense of anticipation. (Sound example 5.2.)
The structure of dispersed and accumulation can be combined to generate more interesting continuation structures. By applying copies of one segment of a dispersed 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-varyingly time-stretching (time-fluctuating, time-varyingly) ebb and flow, we can avoid a merely time-cyclical ebb and flow. More significantly, the closer we cut to the onset of the original sound, the more spectral momenta our accumulation will gather before releasing into the dispersed phase. Hence we can build up a set of related events with different degrees of musical intensity or tension as the accumulative 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 fins too close to the liver is eaten. Some offcidelines, out of bravado, ask for the fish to be cut within half a breath of the liver, something quite beyond consequences. Sound composition is, fortunately, a little less dangerous. (See example 5.3).

Again, time-varying, spatial motion or different types of spectral mislocations (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 more obvious examples of these are: undulation of pitch (vibrato) and undulation of loudness (breathings). Undulating continuation is related to physical activities like shaking and is no accident that a wide, till-like, vibrato is known as a "shaker". 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 (7). This may also be induced in elastic physical objects (like thin metal sheets, this wooden board etc.) by physically shaking them. (See example 5.4).

In naturally occurring vibrato and tremolando, there is moment-to-moment instability of undulating speed (frequency) and undulation depth (pitch-excision for vibrato, loudness fluctuation for tremolando) which is not immediately obvious until we create artificial vibrato or tremolo in which these fluctuations are completely regular. Completely regular speed, in particular, gives the undulation a cyclical, or rhythmic quality drawing our attention to its rhythmicity. (See 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 E-flat or E-flat is added, a common phenomenon. (See example 5.6).

Vocal vibrato is in fact a complex phenomenon. Although the pitch and therefore the partials of the sound shift up and down in frequency, for a given vowel within the spectral peaks (formants) remain where they are, or, in diaphons, move independently. (See Appendix 8i and 8b). 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 lifetime in the relatively amplified environment of a spectral peak. Hence, vibrato can be used to add volume to the vocal sound. (See Diagram 3).
definition, there is no smaller time-frame over which it can be meaningfully perceived. This is part of the
definition of a mathematically irrational quantity. Hence, by definition, we cannot meaningfully perceive a
Golden Section, though again we may comparatively perceive that it lies within the bounds of 1 to
1/2(86th–25fth).

I am happy to accept that Fibonacci-derived ratios give a satisfactory comparative percept of 1 to
1/2(86th–25fth) (possibility even a slightly narrower range) but not that, at larger time-frames, there is any
measurable perception involved. In larger 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 inaction becomes.

I am certain of two things. Firstly that when psycho-acousticians make 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 target of a modern musical mathematician.

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 determinisitic way, i.e. specifying exact time-placement,
exact pitch-value, exact spectral content etc.

If we demand composer 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
organization 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 musical confluence 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 is 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 wavelike 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 is simply dishonest. The composer must then 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 systematics are not, on their own, a guarantee of perceptual
relatedness.
LARGE-DURATION REFERENCE FRAME - 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 quarterbar time-frame) order-sequences are put in motion (rhythm) while at the same time establishing a time-frame (e.g. accented bars) in which longer duration order sequences can be set up.

Within the time-frames fast-denkzeitrufer to 1-every-2-unities, hierarchical systems of interlocked levels functioning as order-sequences in one direction and direction 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 time-scale limit to measured perception. Though many composers writing musical scores would make their approach clear 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 sections) 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 mathmatical advantage of being scalable at longer and longer time-frames. (Diagram 17).

This makes them attractive for work on integral time units, e.g. measures at the fundamental setting of a piece. The question is, can we perceive these proportions in a measured sense, or do we perceive them in a comparative sense (1 to 2; 2 to 3; 3 to 5) while the exact measure makes the task of laying out the score possible. If we consider the golden sections tight (the limit of the Fibonacci series) this will work 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) it's parts are not integrally divisible by any member of fixed polces, no matter how small. Hence it is intrinsically problematic to attempt to score in exact golden sections. There is, of course, such intrinsic problem to scoring tap-tap-segments to such irrational proportions and hence, hypothetically, a perfect golden section setting might be achieved in a musical composition. Whether we are reasonably perceiving this handbook is quite a different question.

On the relatively small-time-frame of 'swing' (see earlier discussion), subtle but comparatively perception of proportions may be very significant (just as there are very subtle parameters in the special character of grain which we cannot 'hear out' but which are fundamental to our neuro-physiological perception). However, on much larger time-frames, we run into the problems of both longer term time memory and the influence of smaller scale event patterns on our sense of the passage of time (swinging ten minnares for this music scored an interesting 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 1/3 can be found within it and the nearer these proportions are to the Golden Ratio the more likely they are to occur in the series. The question is, do we believe that we can reasonably perceive these proportions (e.g. 34:21) or merely that we can comparatively perceive them all as lying within the range 1 to 128th-24ths. More dramatically, because the Golden Section is an irrational quantity, then, by

Ideally we should separate format frames before adding some new breath property to a sound, reorganizing the structure motion of the elements on the pitch-vector sound. (See format preserving spectral manipulation: Appendix p37). In practice we can often get away with more time-speed variation transpositions of the original sound source.

We may extract the ongoing evolution of sound properties using pitch-mapping (Appendix p49-50) or envelope following (Appendix p58) and apply the extracted data to other events. We may also modify (e.g. expanding, compression) : Appendix p46 the unfolding properties of the source-sound. Alternatively, we may derive modifications in pitch (vowels) or intonation (impressions) in a sound, with control over time-varying speed and depth. In this way we may, e.g. produce voice-like articulations on sounds with static spectra, or non-musical environmental sounds (e.g.the sound of a power drill). (Sound example 5.7).

We may also produce extreme, or non-naturally occurring, articulations using extremely wide (more or less) vibrato, or pitch (involvements) to the point of sound granulation. At all stages the overall 'trajectory' (pitch variation of speed and depth) will be important compositional parameters. (Sound example 5.8).

In the limit (very wide and slow) granulation becomes large-scale busheus linearity and we may progress from unifying continuous to fused articulation (and vice versa). Similarly, vibrato (very wide and slow) becomes the fused articulation of moving pitch. We thus move from the unifying articulation of a stable Epoch to the limits of pitch motion structure in which Epoch has no further significance. (Sound example 5.9).

At the opposite extreme, tremolo 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 intermittent aggregation and this may all happen within the ongoing flow of a single event. (Sound example 8.10).

We may also imagine oscillatory continuation of a sound's spectrum, fluctuating in its harmonic/stimmetry dimension, its Stalin-seompe dimension, or as a format-position dimension (spectral unification). Formural modification like hard-'yold'ing, head-shake flavors or "synnex" unification) can be produced and controlled entirely vocally. (Sound example 8.11).

FORCED CONTINUATION AND ITS DEVELOPMENT

In any system where the energizing source has to be continually applied to sustain the sound (e.g. bowed string, blown reed, speech) the activator exerts continuous control over the evolution of the sound. With an ensemble, a player can force a particular type of continuation, a crescendo, a vibrato or tremolo, or an overblown 'quiety' 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 color generation as in a string or wind machine) will tell us how much energy is being applied to the sounding system. These forced tremolo, spectral or pitch movement stages may thus be thought of as physical gestures translated into sound.

Any general 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
Crested) by applying an appropriate loudness trajectory (enveloping) with perhaps more subtle parallelizing features (spatial enhancement by filtering, homogeneity shifting by spectral reshaping or subtle delay). Moreover, these subtly crested shapes can transmute the boundaries of the physically likely C. Sound which is necessarily quiet in the real world (proposed whispering) can be unreasonably loud, while sounds we associate with great forbiddences, e.g. the clanging together of large, heavy objects, the forced paining 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). Mapping it to the original sound (envelope subtraction), or transforming it to other sounds (enveloping or envelope substitution: see Appendix 8).

Information can be extracted from instrumental p=quurance (which we might specifically compose or improvise for the purpose), speech or vocal improvisation (we also from fusing unpredictable phenomena (the dropping of a tap) or working in stereo, the character of a whole field of naturally occurring events (e.g. traffic flow, the crowning of bows etc.). The extracted temporal information can then be modified and supplied to the same material (envelope transformation followed by envelope subtraction), or applied to some entirely different musical phenomena (enveloping or envelope substitution: see Appendix 8) or sound 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 movements 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, grime-fumes 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 indistinguishable quality of grain or of sound mass. (Sound example 8.12).

The movement from mono into stereo may also play a significant role in dynamic incursion (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 (sign(drainage, granular-essence and other types of brassage).

The most obvious way to create a continuation is through time-stretching. There are several ways to do this (brassage/scummen, spectral stretching, warped 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 entity can be made to reveal a surprising morphology. This
In our natural example, the mutual-tone frequency difference is 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 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 idealized performances more 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 is a 13-in-the-time-of-1 grouping, and the 6th unit is an 11-in-the-time-of-one grouping, over the same crochet, at [crochet = 120] are only 1/13th of a crochet, or 1/25th of a second (less than 4 milliseconds) apart. It only needs one of their units to be mislabeled for 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 13).

Hence we are not, here, composed a sequence imminently tied to an "exact" notation. The exact notation in fact specifies a class of sequences with similar properties within a firmly definable range. We are in some sense 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 (possible variable) randomization 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 composer domain, the density approach may be more appropriate if we are really looking for the same class of percept 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 "emotionally" related groupings, even the intrinsic "instability" of live performance, we should be aware of the repetition of the bar length or larger-time-frame unit (which we are dividing). We bear a measured way at the larger time-frame level, but we bear only comparatively at smaller time-frames. On the other hand, with computer-processing in generating the sound sequence, we may be able to perceive a very precise "density flow" in the combination of these streams, at least if they are repeated a sufficient number of times. (Sampled note 9.9).

Our comments about notated music are further reinforced if we now organize the material internally so as to contrast 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 mutal pattern reinforcement, larger time-frame reference-frames will not be perceptually established. (Diagram 15).

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

Process also can change a rapid continuation into something in a phrase-time-frame, for example, format-gliding consonant ("C"), ("Y" in English) become slow format transformations ("oo" → "ah" and "T" to "eh"). Conversely, a continuation structure can be locked into the indescribable qualitative character of a grain, through extreme-time-contraction. (Sampled example 8.13).

Continuation may also be generated through reverberation. This principle is used in many acoustic instruments where a sound box allows the energy from an instrument 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. (Sampled example 8.14).

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 delays (due to the reflection process taking place on different physical types of surface) variation of the sound being mixed with the direct sound as it reaches the ear. Such processes can be electronically distinguished. The electronic memory has the added advantage that we can specify the dimensions of similarity or impossible spaces (e.g. infinite reverberation from an "infinitely long" hall, the inside of a barrel 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 generalized continuation properties of our whole sound set. But it can also be used in a specific way to retard or alter the nature of the individual sound events, extending (elements of) the spectrum in time.

Reverberation cannot, however, by itself, extend any unstable or forced continuation properties of a sound. On the contrary it will lengthen these. Pitching pitch will be extended as a pitch band, because variations will simply be averaged in the new waveform. We can, of course, post-hoc, add unstable or forced continuation properties (slurred, swooped, encroaching) to a reverberation extended sound. These will in fact help to subtly the percept continuance-reverberate as a single sound-event. The reverberation part of the sound will appear to be part of the sound production itself, rather than a supplement of a sound box or characteristic of a room. (Sampled example 8.15).

Institute-reverberate models have in fact been used recently to build synthetic models of acoustic instruments from vibrators, where the separation might seem natural. To this.

CONSTRUCTED CONTINUATION : ZIGZAGGING & BRASSAGE

Continuation can be generated by the process of zigzagging. Here, the sound is read in a forward-backward-forward sequence, each reverse-reading staring 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 the end, the sound will appear to be artificially extended. In the simplest cases, the sound may oscillate between two fixed tones in the middle of the sound until told to proceed to the end. This "alternating loop" technique was used in early commercial
samples 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 intensity. In general, however, any segment of a sound, beyond grain time-frame, will have a discontinuous structure and hopping over it will cause this structure to be heard as a mechanical repetition in the sound. (See Appendix p. 44.) (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 thus use the process to achieve spectral, pitch or loudness movements or discontinuities in a sound and focus attention on them by alternating repetition. By shifting the zigzag points slowly from zig to zag, 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 (in a longer timescale) dispersive-acumulative (as shown) effects, within a sound-continuation. (Sound example 5.17)

Sounds can also be artificially confined by bounceage. In the bounceage process successive segments are cut from a sound and then replaced 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 pp.4-5). (see note at end of chapter).

As discussed previously, grain time-frame segments will produce a simple time-stretching of the sound (homogeneous effect). Slightly larger segments may introduce a groove-like percept into the sound as the perceived evolving shapes of the segments are boosted at repeated. (Sound example 5.18). The segment granulation of an already groove-like source may produce unexpected phasing or delay effects within the sound. Longer segments, especially when operating on a segmented source (a groove source) will result in a systematic collapse of its elements. With regular segment-size our intention will be to 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 rhythmic, collapse type excitation. (Sound example 5.19).

This idea of bounceage can be generalized. Using non-regional 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 articulate (brief but longer time-frame) aura to the time-stretched sources. We may also permit the process to select segments from within a time-range measured backwards from the current position to the source sound (see Appendix p. 42). In this way, echo percepts are randomized further. Subtle controlling this and the previous factors, we can extract rich fields of possibilities from the small features of sound with evolving spectra (especially, sequences, which present us with constancy 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 p. 43). 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 bounce in this way using a relatively large segments-size (including several notes) we will create a new rhythmic strain including more and more of the notes in the original melody. The original melody will thus control the evolving Rhythmic field of pitch possibilities in the goal sound. (Sound example 5.21). Or a smaller time-frame, the qualities of a highly characteristic onsets event can be reintroduced over the entire ensuing continuation. (Sound example 5.22).

53
The brussage 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 sound a set of pitches in a very small band providing a stable "stepped" texture in the continuation (particularly if grain-size is slightly random-varied to avoid rhythmic constraints 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 spatialize, or progressively spatialize, the segregated components, moving from a point source to a spread source. (Sound example 5.26.)

Eventually, such evolved manipulations (and their combinations) form a continuous or coherently segregated source to disintegrate 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 generalize the brussage concept further, taking us into the realm of granular reconstruction. As with brussage, our source sound is cut into segments which we may vary in duration, pitch or loudness. However, instead of merely replicating these tail-to-tail, they are used as the elements of an evolving texture in which we can control the density and the rate-variation of the elements. (See Appendix 525.) (Sound example 5.27.)

In this way we can overlay segments in the ensuing stream, or, if segments are very short, introduce momentary silences between the grains. This process, especially when 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 texture characteristics (density, pitch spread, loudness spread, spatial 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 generally thickened (and hence more heavy, or at least less focused), 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 space (see Chapter 8). (Sound example 5.29.)

Final note: in order to simplify the discussion in this chapter and also in the diagrammatic appendix, brussage has been described here as a process in which the cut segments are not overlapped (apart from the length of the edit itself) when they are reassembled. In practice, normal brussage/harmonizer processes use a certain degree of overlap of the adjacent segments to ensure sonic continuity and avoid granular artifacts (where these are not required) in the resulting sound.
This problem becomes particularly important where different divisions of a pulse are superimposed. To give an example, if we compose two duration streams in the proportion 2:3 (see Diagram 8) we have a clear mutual time-frame pulse at the larger time-frame of the crochet. (Read example 9.6).

If we now reorganize the elements in each stream in a way which contradicts this mutual pulse (see Diagram 46) we may still be able to perceptually integrate the streams (perceive their exact relationships) 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). (Read example 5.7).

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. (Read 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. 100ths 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 crochet beat into (say) 7, 1, 1, & 13 at (crochet = 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 realized in practice?

No human performance is strictly rhythmically regular (see discussion of quantized rhythm above). When these streams are laid together the deviations from regularity will not flow in parallel (they will not synchronize like the parallel micro-asynchronies 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 units, 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).
CHAPTER 6

DISCONTINUOUS SPECTRA

In this chapter and the next we will discuss the particular properties of sounds with perceptible 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 occur cyclically), rather than as a sequence of changing but definite sound events. Once, however, the individual spectral events 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, consecutively or inconsecutively. We are here concerned with perceived discontinuous changes in the time range speed-of-normal-speech down to the lower limit of grain perception.

Compositionaliy, we tend to demand different things of discontinuous sounds, depending on their type. In particular, if we time-stretch a continuous sound, we may be dissatisfied by the overall distortion but the remainder of the sound may appear perfectly satisfactory. If we time-stretch a discontinuous sound, however, we will be disappointed everywhere by some distortion as the sound is a sequence of events. Often when we want the sound (e.g. in use-real environment, a drum roll, a speech-simulant) to be delivered more slowly to us we want the individual attacks (the drum strike, the timbral structure of consonants) being answered out and hence transformed. We wish to be additive 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 tempos of the event stream 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 when this (near) ideal limit is approached in naturally occurring sounds (e.g. bow consonation noises) 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 those in the next chapter.

Both grain-streams and sequences can have (or can be composed to have) overall continuation properties (digression, redundancy and found 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 reel or roll on drums, keyed percussion or any sounding material. They are produced vocally by rolled "y" sounds of various sorts in various languages, by lip-flaps and by inhaled caps. Vocally, such sounds may be used to modulate others (i.e., rolled "y", whistled rolled "r", flute-toned woodwind and brass etc). The rapid opening and closing of a resonant cavity containing a sounding source (e.g., the hand over the mouth as we sing or whistle) can be used to naturally grain-stream any sound. (Sampled 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 envelop following another sound (see Chapter 10). (On-off) might be something from a deep shuffling fluttering to an almost on-off gating of the signal. A particular element is to achieve the effect to generate a loudness trajectory on a sound tied to the wavelets or wave-sequences it contains (wavelet enveloping). If each on-off type trajectory is between about 25 and 100 wavelets is length, we hear grain-streaming. (Below this limit, we may produce a rasping or special "cursing" of the source sound.) This process des the grain-streaming to the internal properties of the source sound as, for example, the grain- streaming standards if the perceived pitch falls. This suggests to the ear that passing postamentos and the standard are causally linked and intrinsic to the goal-sound, rather than a compensatory affixation (1). (Sampled example 6.4.)

Alternatively grain-streams may be constructed by altering together individual grains (1). 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 (distortion) produces a more convincingly natural result (Sampled example 6.5).

More compositionally flexible, but more procedurally, is to use a zoning program on the individual grains to be placed and ordered, then repositioned, replaced or reordered using meta-instruments which allow us to manipulate mixing instructions (enveloping) or to generate and manipulate time-sequences (sequence generation).

In this way grain-streams can be given gestural acoustics or standards of different shapes and be slightly randomized to avoid a mechanistic result. (Sampled 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 wavelet distortion, wavelet harmonics distortion or warlow—see Chapter 3) combined perhaps with inherrowning (see Chapter 12) to generate a set of intermediate sound-states. (Sampled example 6.7.)

Short continuous sounds can be extended into longer grain-streamed sound by using bourgeoisie with appropriate grain time-frame (duration) segment-size. (Sampled example 6.8.) The granulation of the resulting sound can be exaggerated by convolution (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 aware of continuous sounds with grain-streams in this way and hence begin to build networks or musical relationships amongst diverse materials.

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 mensural reference-frame, and the sequence is played "precisely" as written, we will perceive the 3:1:2:1:3:1 etc. sequence of durational 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 perceived by the performer, veering towards 2:1:2:1:3, 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 gives an illusionary to 3:1 2:1:3, but we are concerned here with the percept. (Sampled example 9.4.)

The way in which such crotchet beats are divided in one aspect of a sense of "swing" in certain styles of music. A particular drumbeat, for example, may have at its foundation the regularity of the drum-beat and the articulation of the particular quality of this division in the sense of the particular sense of swing it imparts to both player and remaining completely unaware of the exact numerical proportions involved. Here, then we have a comparative perception with fundamental qualitative consequences. (Sampled 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 to some extent normal in the extent that e.g., the precision morphology of each violin note cannot be specified in the manner and varies arbitrarily over a small range of possibilities, i.e., the drumbeat on open staves. In sound composition, these factors can be precisely specified and, if desired, reduced to the level of comparative, or measured, perception.

More importantly, our rhythmic example using divided 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 "independence factors" proposed by Clarence Barlow (see Chapter 7). As discussed previously, we can define, over a reference frame of quavers, the relative independency of each note in a 1/4 or 1/2 (or 1/8) pattern. Linking the possibility of occurrence of a note to its independency factor allows us to generate a strong 1/4 grouping, or a 3/4 grouping, or an ambiguous percept between the two. Once every group becomes equally probable, however, the sense of grouping breaks down altogether and at the level of 1/6 groupings (1/6-groupings, or any groupings) measured perception is lost. Our perception returns 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 swing phenomena. On the large scale, if the pulse becomes too long (e.g., 50 notes), 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.

56

75
DIAGRAM 1

DIAGRAM 2

DIAGRAM 3

DIAGRAM 4

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 wave spread-overlap or spectral time-overlap (with no grain pitch alteration) may force the grain separation under an optimum time limit for grain-perception and the granular 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-stretching: see below) we will breach the grain-perception limit. The sound will gradually become a continuous flow. In this case a related pitch may or may not emerge. (Sound example 6.10) Reversing 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 discontinuous in other ways...

1. skewing 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. skewing down the sequence and the grains, so the internal morphology of the individual events comes to the fore (granular stretch-overlap: Sound example 6.14).

3. gradually shifting the pitch or spectral quality of different grains differently, so the grain-stream becomes a sequence (granular rounding: 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 timing 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 dealt with note events. By the appropriate use of gating, cutting and delaying or rounding, which may be all combined in a single sound processing instrument, we can refine the grains in a sequence using various mathematical templates (slow by a fixed factor, resembling arithmetically, geometrically or exponentially, randomised grain locations about their mean, stretch the time-frame in similar ways and to create granular time-variations. The grain-stream can then be elegantly time-dissolved without discarding the grain-segments. (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 reverser would recognize 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).
MEASURED, COMPARATIVE & TEXTURAL PERCEPTION

To make these distinctions completely clear we must first ascertain the nature of time–order perception and clarify our terms more precisely.

In the ensuing discussion we will use the term ‘duration’ to mean events–must–separations–duration—to make things easier to read 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 these types—the overheard, the comparative and the intended (the perception of temporal identity and density relationship).

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 organization of a piece so that at one level (e.g. a metrical or phrase time–frame) order–sequences are set in motion (rhythm) while simultaneously establishing a time–frame at a 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 intended?

The fine assertion we will make is that measured perception requires an established or only slowly changing reference–frame against which we can ‘measure’ when 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 F–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, quarter notes at (bar 130) we can recognize with a fair degree of precision the various multiples of dotted crotchets, minims, etc. and integral divisions (semiminquavers, demiminquavers, triplet semiquavers) of this unit. (See Diagram 1). Hence we can, in many cases, measure our perception against the reference–frame and recognize 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 durational sequence of duration values in the perceivable measured ratio 1:2:3:4. 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 perception of ‘flow–ing down’ but we do not (normally) have a clear perception of the exact proportions appearing between the successive events. Furthermore no two live performances of the events will preserve the same set of measured proportions between the consistencies. This is, then, an example of comparative perception.
In this sequence we probably retain the sense of an underlying measured pace (which is being articulated by random scattering) to 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 bearing the reference frame, we would perhaps be more willing to declare them arrhythmic.

Taking the sequence in the opposite order, there may be a presuppositional switching point at which we suddenly become aware of rhythmic order in the sequence. (Island 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-sequences 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-quantization grid). Allowing event-sequences 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 streams. 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 Hilpich reference frame. Here we would begin with events confined to an Hilpich set (a Harmonic field), then gradually randomize the timing of some from the Hilpich set, slowly destroying the Hilpich field characterization, 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 it an ordering relation over such a reference frame. Dissolving rhythmicity in terms analogous to dissolving the percept of Hilpich, which also retains synchronically 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 semitones for scales played in the Western tempered scale, the octaves 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 isolated form, this may also be thought of as the time-quantization 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, group-level and continuation-level perception (or lack of it) into longer swathes of time. Moreover, because such time-frames may be nested (as 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 Hilpich perception, it is the dissolution of the time-frame which leads us from field-ordered (rhythmic) perception of temporal organization to density perception. And just as dissolving the Hilpich percept by the randomization of timing 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-organization discussed in the previous chapter.
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 sounds of distinct kinds of birdsong, klangfarbenmelodien passing between the instruments of an ensemble, a "rundum" on a multi-instrument percussion set etc. We can also construct disjoint sequences of arbitrary sounds by simply splicing them together (e.g. a dripping tap, a car horn, an obese note, a cough - with environmentally appropriate or inappropriate balance and/or reverberation relations one to the other) or by modifying existing natural sequences (time-conversion of speech or music, for example).
(Sound example 7.1)

Naturally occurring sequence 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 (1944) 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 notice in our writing system - rather than those elements themselves (this is Diphasic synthesis). Starting from the separate elements themselves, to achieve the flowing entity of each natural prompts to 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 or 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 & DESTROYING SEQUENCES

Sequences can be generated in many ways, apart from splicing together the individual segments "by hand". Any sound source in directed motion (e.g. a pitch-gliss or a. former glide) can be spliced into elements which, when rearranged, do not mean the spectral-continuity of the original. (Sound example 7.2)

An existing speech-stream can be similarly rendered to the synthetic system and (depending on where we cut) the phoneme-continuity. We may do this by chopping up the sequence into contiguous segments and reordering them (as in sound shuffling - see below), or by selecting segments in some random order (to which they may overlap other chosen segments) and reassembling them as they are spliced back together again (random cutting - Appendix p41). (Sound example 7.2)

We may work with a definite sequence 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, roughly each in a non-progressive manner (different filterings, pitch shifts, etc) and reassemble the original sequence by replicating the elements together again, but they will now have discontinuously varying properties. (Sound example 7.4)

ABOUT TIME

In this chapter we will discuss various aspects of the organisation of time in sound perceptions. 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-sequences which were temporally dense but which might remain field properties in other dimensions (like pitch or format-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 conluded over a given range, a pitch density value would tell us how densely the pitch-events cover the combined 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 concept of Density and Field applied over the same parameter come into conflict. Once the pitch-density (in this new sense) becomes very high, we lose any sense of a specific pitch field or Identifiable reference frame, though we may continue to be aware of the range limits of the field. (Sound example 8.1).

We must therefore ask, 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. pitch organisation).

Compositionaly, we can create sequences of events that gradually lose a (measured) sense of rhythm. Thus we might begin with a computer-generated rhythmic sequence which has "natural" or "inarticulate" precision. Adding a very small amount of random scatter to the time-position of the events, given the rhythm is 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 "stuttering" 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 rhythms, or even badly performed rhythms and eventually the rhythm percept is lost. The time-sequencing is artificial. Once this point is reached we perceive the event succession as having a certain density of event events - our perception has changed from grasping rhythmic ordering as such to grasping only density and density fluctuations. (Sound example 9.2).
Using Knowske techniques, we may achieve similar results, if the sequence size is not notably large and
we work on a clearly spectrally-evolving source. Using several browsers 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 spectral in the order sequence
of the source but, in the meantime, retain unpredictably already revealed materials. (Sound example
7.8).

Browser with pitch variations in a finite pitch set (possibly cyclical) could establish an
Hitchcock—sequence from a pitch—continuum (or other source). (Sound example 7.6). Even browsers
with spatialization (see Appendix p48) will separate a continuity (or any other) source into a spatial
sequence and so long as they are spatially determined, 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 reversed 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 circumstances of a sequence of
regularly spaced events with strong attacks, a pitch peak. Conversely, on—averaging the sequence beyond a
certain limit will bring the internal properties of the sequence into the planes time—frame and the
sequence of events will become a formal property of the large time—frame. (Sound example 7.8).

Alternatively, copies of the sequence, or groups of its continuums may be inserted in texture (see
Chapter 8) of sufficient density that we 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 sequences which are then reordered randomly. This process may be repeated
to infinitum, gradually decaying all but the most persistent spectral properties of the sequence to a
warty complexity (the complex in fact becomes a simple tachytonic process). (See Appendix p41). (Sound
example 7.10).

GENERAL PROPERTIES OF SEQUENCES

All sequences have two general properties. They are defined both a field and an order. Thus, on the
large scale, the set of structures for a particular natural language defines a field, the set of phonemes
which are used in the language. Similarly any sequence played on a piano defines the tuning set or
fundamental 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 an Hitchcock 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 contains. (Sound example 7.11).

Sequences are also characterized by order properties. In existing musical languages, certain sequences
of notes will be commonplace, others exceedingly unlikely. In a particular natural language certain
clustering of consonants (e.g., "not" in English) will be commonplace, others rare; and yet others
absent. It is easy to imagine and to generate meaningful sequences, though the human mind's pattern
seeking predilections makes it over—deterministic, hearing definite patterns where patterns are 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 seen as musically competent elements. On the larger scale, these may be predetermined by cultural norms, like tuning systems or the phonetic 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 relationships of these subsets.

In this context the size of the field is significant. Musical settings of tone, for example, may treat the Hypich (or chromatic) set in terms of a small ordered reference on which is constantly restructured in subsets (e.g. chord formations over a scale) and rearranged (metric variation), whereas the phonetic musical materials in 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 *metrical*.

Poetry, however, through rhyme, alliteration and particularly rhythm, begins to adopt the small-scale reference-frame and order sequencing for the phenomena we find normal in traditional musical practice. We therefore draw-on a meeting ground between phonetic and traditional Hypich and *metrical* as well as and these connections have been explored by sound poets (Antonioli, etc) and composers (Noyo, etc) alike. As we move towards poetry which is more strongly focused in the somatics of words, as just as of syllabic utterance, the importance of small-scale reference-frame and order sequencing may become overshadowing (e.g. Schwitter, Uranium). (Sound example 7.12.

COMPOSING FIELD PROPERTIES

Constructing sequences from existing non-sequence, or differently sequenced, objects (a flute melody, or an upward sweeping tone-event, or a traffic recording in a tunnel with a particular noise resonance... or a conversation in Japanese) means that some field properties of the source sounds will enter into the resulting sequence: a defined reference set and a flute spectrum (with or without other characteristics); in the latter case the field is altered, using noise-bands within a given range, the resonance characteristics of the source, the spectral characteristics of the phonemes of Japanese and perhaps the set 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) 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 distorsion with subharmonics, see Chapter 13) 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 reverboration, delay, small-time-frame language, spectral blurring, spectral sharpening, waveform sharpening, or granular (or 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

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

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

In Sound example 8.15 (from Red Bird: 1977, a pre-digital work), a noisy-white noise is superimposed with a modified skylark song via this process of temporal thickening. In Sound example 8.16 (from Vos 5: 1986), behind the stream of complex vocal multiphonics, we hear the sound of a crowd emerging from a vocal-like noise band which in itself is 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 Brazil it is sometimes possible to become completely snow-covered. If the snow fall is insufficiently heavy, some are left alone in the landscape in the watter of white snow particles. Similarly, when a sound texture is made extremely dense, we lose perceptual track of the individual elements and the sound becomes more and more like a continuum. In particular, when the sound elements are of many special 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 white-out and the resulting noise-band is then filtered to give it spectral pitches while it slided up and down. The pitch 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 retain 'contradictory' properties. Thus a texture may consist of events whose causts are entirely randomly distributed in time but whose event-elements are clearly rhythmically ordered within themselves, or, conversely, events may begin on the pitch of a clearly defined harmonic field while the event-elements have harmonics distributed randomly over the continuum. (Sound example 8.18).
We cannot hope to describe all possible conceivable 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 (harmonic—inharmonic; type of idiosyncrasy; format-type; resistance—stability; stability of 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 spatialization of groups (moving left, spatial oscillation, spatial randomness), range of spatialization types, and the time-variation of these.

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

All such features may be comparatively 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 increase and decrease 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. (See 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 impetuses 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 elastics. Spatial variation in temperature similarly too only be measured over an even larger mass — and temporal variation in density similarly requires more events, more time, to be perceived than 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 random—frame. Furthermore, slow density fluctuations will tend to be perceived as directed flows — as continuation properties — as compared with towards a simple continuation trend. In these ways we can form bridges between sequences, grain streams and smoothly continuous sound, establishing audible links between musically diverse materials. (See example 7.15.)

We may also focus the field—and—properties of a sequence or part of a sequence by looping (repeating that portion). Thus a series of pitches may become salient, and may in fact be entirely removed across the continuum, establishing no reference frame of Pitch(s). If, however, we repeat any small subset of the sequence, it will very rapidly establish itself in its own reference set. We can then move very rapidly from a sense of Pitch inherent to Pitch definitions. (See example 7.16.)

The same set—focusing phenomena applies to the spectral characteristics of sounds. If a meaningful group of sounds is repeated over and over again we eventually lose contact with meaning in the phrase as we become aware than the material is just a sequence of some events defining a sonic reference frame. (See 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—set of the manipulation of finite sets of entities which has been very extensively researched and used in last Twentieth Century serial composition. Starting originally as a method for organizing pitches which have specific relational properties due to the cyclical nature of the pitch domain (see Chapter 11) it was expanded fast to the time—domain, which has some similar properties, and then to all aspects of related compositions.

Permutation of a finite set of elements in fact provide 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 (citations that they are perceptually inter—related and often open to disassemble). 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 in—ordering relation — or as an underlying field construct?

In traditional serial music, the manipulation of set—rule relations is usually confined to a subset of properties of the sounds e.g. Pitch, fixed duration values, a set of timbral values. But we can work with alternative properties, or groups of properties, such as in total spectral type. For example, permutations of instrument type in a monodic long/short/multi—voice — Webern’s Opus 21 Saluté provides us with such an instrument-to—instrument line and more complexly. We could, however, perceive the instrumental sequence and not the Hypothes. Formal sequence e.g.

\[ ba – ba = ba – ba = ba – ba \]

and some characteristic sequence e.g.

\[ ba – ba = ba – ba = ba – ba \]
can be most easily observed in patronal properties in language or language-derived syllables, but formal and onset characteristics can be extricated and attended to 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", "dark", "soft", "stiffened" etc. attacks as a definable sequence. The combinations of such sounds might be recognizable (e.g. words, speech, a bell, a piano concerto except re) but ordering principles are applied only as 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 in contrast the area of order manipulation with that of lesser order. In the setting of prose this contrasts is clear. In our extreme example we have suggested a kind of ordering of onset characteristics built against a "cinematographic" message which might have an alternative narrative logic – the two might offset and comment upon one another in the same way that metric 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-ring where the sequence in which a fixed set of bells is rung goes through an ordered sequence of permutations. (See Diagram 1.) (Bound example 7.18.)

STRESS PATTERNS AND RHYTHM

We can apply different order-groupings to sequentially exclusive sets of properties (e.g. count, continuation, loudness, as against Hitchcock within the same reference-set of sounds. Onset characteristics and overall loudness are often organized cooperatively to create stress patterns in sequences. Combined with the organization of durations, we have the group of properties used to define rhythm. Rhythmic structure can be established independently of pitch structure (as in a drums-alone solo) or parallel to and independently of Hitchcock order relations. Most of the 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-independently but mutually interactive way (Harmonic rhythm in Western Music, meters/rhythmic cadence in Indian Classical Music).

Again, whole volumes might be devoted to rhythmic organization 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 organization, or on a regular duration pattern that is subsequently permuted or time-warped in a consciously varying manner, a feature over which we can have complete control in sound composition. Hence, stress duration patterns at different tempos, or at different varying tempos, can be organized in such a way that they synchronize or largely defined lines, these lines perhaps themselves governed by larger time-frame stress patterns. (See Chapter 9.) (Bound 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 groupings perceptions on the same stream of sema-paucity units, e.g. grouped in 5 and in 7 and in 4 and in 4 and in 4 and to change their relative perceptual prominence by subtly altering loudness balance or onset characteristics or in fact by dynamically grouping perception by a gradual randomization of stress information. Here field variation = adding on order perception. (Bound example 7.20.)

63
Field properties may also be evolving through time, in a direct or oscillating manner. One method of creating such evolution is through the use of a texture whose elements are gradually changing to one. In Sound Example 8.11 we have a texture of vocal utterances which becomes very dense and then rare. As it does so, its elements are specificaly pitched (see Appendix 2B) so that the texture—stream becomes a 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 initially by surrounding (see Chapter 9) or by some class (e.g., twang traps, pedals on tone, bricks into steel, water drops etc.) then gradually 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 (widening Appendix 2D).

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 crossing the individual Hpitch values towards new values. This kind of gradual crossing shifts 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 now becomes drawn from a constantly increasing range about a medium value. Initially we retain the same 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 dehyster the Hpitch field by gliding the pitches. Once the gliding is large, and not start— or end—bound (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 8.3).
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 in our perception of that grouping. In 6/8 for example, the 6th note (which begins the 2nd set of 3 notes) 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 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-notes versus 3/4-notes. Also, by making all notes equally probable, all notes regular in length and the sequence becomes arhythmic above the quarter pulse level. (Diagram 2). (Ibid. Example 7,21.)

ORDERING ORDER

We can also work with order-groupings of order-groupings e.g. with Hipchises, altering the time-succession of motives-groupings—taking shape. There is an extensive musical literature discussing order-groupings, especially of Hipchises, and all such considerations can be applied to any subset of the properties of sounds, or sounds as wholes.

Here we are beginning to stray beyond the frame of reference of this chapter, because ordering 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 arm economy, and the less easy persistence—reversal 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 "ordered" objects are not increasingly varied in more complicated ways at smaller time-frames, but at large time-blocks qua large time-beats, the persistence structure tends to be simpler. (Such matters are discussed in greater detail in Chapter 4.)

The computer, having no ear or human musical judgment, can manipulate order-sequences of any length, and of any order—size in an entirely equivalent manner. As a composer, however, we must be cleare about the relationships of such operations to our time experience. Equivalence in the numerical domain of the computer (as for that matter on the spatial surface of music) is not the same as experiential equivalence. Audible design requires musical judgment 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 expansion of a sequence (by self-repeating, over—extending etc.) is a way of passing from one perceptual time-frame to another, just as time—shifting will ultimately connect the sequence into an essentially continuous perceptual event. With time—expansion, the perceptual boundaries are less clear, and so less important.

Permutational logic is heady stuff and relationships of structure can breed in an explosive fashion, like Fibonacci's rabbits. Permuting and otherwise rearranging sequences in even more complicated ways is something that computers do with consummate ease. A powerful sound—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 reworkings and if so, are they heard directly, or as field contrasts on the total

Alternatively we may begin with individually longer sound sources. Grain masses and sequences may begin to take on texture—strum characteristics by becoming arhythmic through the scattering of onset—time away from any time reference—frame; we take in this Chapter and be by becoming sequentially disordered. If we now begin to superimpose variants, even with a few superimpositions we will certainly generate a texture stream. (Example 8.7)

Continuous sounds which are evolving through time (pitch—glide, spectral change etc) may also change into texture—strum. Applying granular recomposition (see Appendix II) 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. (Example 8.8)

Specialisation 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 splits 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—strum 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. (Example 8.9."

And in fact any dense and complex sequence of, possibly layered, musical events, originally having, sequential, compositional or other logical, 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 as being present over a whole—time scale, even though we do not receive the exact sequence of Hipchises. In this use, we mean a rhythmic concept. 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, syllable—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 agglomeration over a rhythmic mass, but the placement is quantised over a very rigid pulse (e.g. events only occur on time—onset at multiples of 1/100th of a second), we may still be aware of that regular time—grain in the texture. This is a field percept. A more complete discussion of temporal perception can be found in the next chapter. (Example 8.10)
So, broadly speaking, texture is sequence in which no order is perceived, whether or not order is intended or defensible in any mathematical or notatable way. Texture differs from Continuo in that we retain a sense that the sound event is composed of many distinct events. Pure sequential 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 undoubtedly 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 one.

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

Alternatively textual elements may be submitted as ‘samples’ (or a ‘sampler’) or, equivalently, as sampled sound in a look-up table for substitution to a table-reading instrument like Closed. 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 Closed (textile) score.

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

Via such texture control or texture generation procedures (see Appendix pp.48-49), we can generate a texture for a specified duration from any number of sound-sources, giving information on temporal-density of events, degree of randomization of onset-time, type (or absence) of event—event quantitation (the smallest unit of the time-grid on which events must be located), pitch-ranges (defined over the continuum or a preselected, possibly time-varying, Harmonic field), range of loudness from which each event gets a specific loudness, individual event durations, 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 pp.48-49). (Sound example 8.5).

Thirdly, any shortish sound may be used as the basis for a texture—stream, if used as an element in generational rulesets where the onset time distribution is randomized and the density is high, but not extremely high. The properties of the sound (loudness irregularity, spectral brightness etc.) may be varied from unit to unit to give a diversity of texture—stream elements. (Sound example 8.6).
experiences? If the latter, need they be quite so involved as I suppose, i.e. in there a simple, and hence more elegant, way to achieve an equivalent mental experience? A more difficult question, as with any musical process is, does it matter to the listener that they hear this modeling process? Or, what does this modeling "mean" in the context of an entire composition? Is it another way of looking down a kilodiscourse 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 mixed events? These are, of 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 those agglomerations of the same or similar (e.g. tape-speed variations) 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 discontinuity of events in time.

In the first two sound examples, we give what we hope will be (at this stage) indescribable examples of measured perceived sound and texturally perceived sound. In the first we are aware of specific order relations amongst Nyquistches and the relative instant of time of events and can, so to speak, reason them, or assign them specific values (this will be explained more fully later). In the latter we hear a whole 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 vague definable organizing properties - the Nyquistches 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. First, the elements may be relatively disordered (random) in some property (e.g. Nyquistches). Second, 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 impossible to pick up on a local ordering in a totally disordered sequence. Thus a random sequence of notes and orders must contain the sequence 11001100 which we may momentarily perceive as ordered, even if the pattern fails to persist. Such focusing on treatment ordisorderess may be a feature of human perception as we tend to be intransitive 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 ever-new-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 Nyquistches 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.
experience? If the latter, are they as just as such an experience? A more difficult question, as with any musical process, is: does it matter to the listener that they hear this recording process? Or, what does this recording "mean" in the context of an entire composition? Is it another way of looking at 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 music, but, 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 aggregations of the same or similar (e.g., tape-speed variation) 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 analyze in more detail the boundary between musical and perceived musical, 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) indistinguishable examples of measured perceived sound and texturally perceived sound. In the first we are aware of specific order relations amongst Hitches and the relative constant time of events and can, so to speak, measure them, or assign some specific values (this will be explained more fully later). In the latter we hear a webbed 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 definitely existing properties — the Hitchets 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., pitch). 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 a local ordering in a usually unordered sequence. Thus a random sequence of zeros and ones may contain the sequence 100000 which we may momentarily perceive as ordered, even if the pattern falls to justify. Such focusing on transient orderliness may be a feature of human perception as we tend to invent 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 override the effect of event-order 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 Hitches 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 Continuums in that we retain a sense that the sound event is composed of many discrete events. Pure serial 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 seductively that we do not latch onto a particular spectral reference frame and we hear an average concept, "noise". But, like noise, texture consists 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 unreplicated sound-sources, specifying their timing and loudness through a 'singing score' (or a graphic mixing environment), though in high density cases this can be less powerful) and use various mix shuffling meta-instruments to control mixing, loudness and sound-source order.

Alternatively textual elements may be submitted as 'samples' (as a 'sampled') or, equivalently, as sampled sound in a look-up table for substitution in a table-reading instrument like Clonat. 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 MARE parameters we wish to assign to the MIDI output) - or from a Clonat (textile) score.

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

Via rich texture control or texture generation procedures (see Appendix pp.68-69), we can generate a texture for a specified duration from any number of sound-sources, giving information on temporal-density of events, degree of democratization of onset-time-type (or abstract) of event-onset quantization the smallest unit of the time-grid on which events must be located, pitch-range (defined over the continuum or a proscribed, 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--sound. In addition all of these parameters may vary (indiscreetly) through time (see Appendix pp.68-69). (Sound example 8.5).

Thirdly, any sound or sound may be used as the basis for a texture-stream. if used as an element in procedural synthesis where the onset time distribution is random and the density is high, but not extremely high. The properties of the sound (sound trajectory, spectral brightness etc.) may be varied from unit to unit to give a diversity of texture-stream elements. (Sound example 8.6).
Alternatively, Clarence Barlow has defined an "indeterminacy factor" for each event in a rhythmic grouping, e.g., a 6/8 bar. This defines how imprecise the presence or absence of an event is to our perception of that grouping. In 6/8 for example, the 4th note (which begins the 2nd set of 3 notes) is the act, and so on. In 3/4 over the same 6 quavers, the indeterminacy factors will be rearranged in a different order, so redirecting the sequence to the 6/8 group 3 of 2. These features are then used to define the probability that a note will occur in any rhythmic pattern and by varying these probabilities we can e.g., vary our perception of 6/8-mass versus 3/4-mass. Also, by making all notes equally probable, all stress regularity is lost and the sequence becomes indeterminate above the quaver pulse level. (Diagram 2). (Sample example 7.21).

ORDERING ORDER

We can also work with order-groupings of order-groupings e.g., with Hipchips, altering the time-sequence of motif-groupings taken as wholes. There is a large amount of literature concerning order-progressions, especially of Hipchips, and all such considerations can be applied to any aspect of the properties of sounds, or sounds as wholes.

Here we are beginning to stray beyond the frame of reference of this chapter, because recording principles can be applied on larger and larger time-frames and are the reference of traditional musical "time". From a perceptual point of view, it would appear that the larger the time-frame (phrase-frame and beyond), the more preconception is dependent on long term memory, and the less easy pattern-recognition 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-blocks qua large time-blocks, the "repeated" structure tends to become simpler. (Such matters are discussed in greater detail in Chapter 9).

The composer, having no eye 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 dimension of the composer (or for that matter, the performer) is not the same as perceptual equivalence. Audible design requires musical judgement and cannot be "justified" on the basis of the computer-logic of a generating process.

In fact, 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 repeating, figure-repeating etc.) is a way of passing from one perceptual time-frame to another, just as time-skipping will ultimately construct the sequence into an essentially continuous perceptual event. With time-variation, the perceptual boundaries are less clear, but no less important.

Permutations of facts are both useful and relationships of combinations can be heard in an expansive fashion, like Fibonacci's rabbit. Permuting and otherwise recording sequences in ever more complicated ways is something the composers do with commonplace sense. 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 these internalizations and if so, are they heard directly, or as field comments on the total.

Alternatively we may begin with individually longer sound sources. Great streams and sequences may begin to take on texture-seam characteristics by becoming acoustic through the scattering of state-time away from any time-reference-frame; see below 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. (Sample example 8.7).

Continuous sounds which are evolving through time (pitch-glides, spatial change etc.) may also change into texture-textures. Applying granular resummation (see Appendix Fig. 2) to such a continuous stream with a grain-size that is large enough to mean 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. (Sample example 8.8).

Specialisation can be a factor in the emergence of a texture-texture. If a small sequence has its elements distributed alternately between extreme left and right positions, the percept will be much more likely 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-texture concept as the sense of sequential continuity is destroyed, particularly if the event-time distribution is randomised. A superimposition of two such scattered sequences will almost certainly merge into a texture-texture. (Sample example 8.9).

And in fact any dense and complex sequence of, possibly layered, layered events, originally having sequential, encompassual and other logics, can become a texture-texture 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-time scale, even though we do not retain the exact sequence of Hipchips. In this case, we retain a "sensibility" percept. Similarly, we may be able to distinguish French from Portuguese, or Classical from American, even if we do not speak those languages and hence do not latch onto significant sequences in the pitch stream. This is possible because the sound-forms, component-types, symbolic-combinations or even the pitch-visualisation 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 indivisibleagnation over a rhythmic theme, 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. (Sample example 8.10).
Field properties themselves may also be evolving through time, in a direct or oscillating manner. One method of creating sound variability (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 more dense and rising in pitch. As it does so, its elements are specifically-scored (see Appendix E.A) 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 snap, cymbal on titles, bricks into mud, water drops etc.) then a secondarily interpretive (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 (swirling up as Appendix G.0). If a field property has a reference-frame (see Chapter 1), e.g. an Haptic set for pitches, these field properties of the texture may gradually change in a number of distinct ways.

In terms of an Haptic field...

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

(b) We may change the Haptic field by gradually reversing the individual Haptic values towards new values. This kind of gradual meter shift can destroy the initial Haptic concept and reveal the existence of the pitch continuum. (see Diagram 2b).

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

(d) We may gradually change the Haptic field by gliding the pitches. Once the gliding is large, and not start- or end- focused (see Chapter 2), we have the original Haptic concept. (See Diagram 2d).

Each reference-frame variation can be applied to other properties which are perceived against a reference frame. (Sound Example 8.12).
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", "haut", "soft", "shadowed" etc. attacks as a definable sequence. The continuations of such sounds might be recognizable (e.g. waves, speech, a bell, a piano concerto excepts etc.) but ordering principles are applied just to the onset type. This is an extreme example because I wish to stress the point that order can be applied in 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. The setting of order plus context is clear. In our extreme example we have suggested a kind of ordering of onset characteristics laid against a “cognitive” montage which might have an alternative narrative logic – the two might offset and correct upon one another in the same way that music (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 four- or six of sounds can be rearranged as wholes and these order relations explored and observed. A traditional example is English change-ringing where the sequence in which a set of bells is rung goes through an ordered sequence of permutations (see Diagram 1). (See example 7.16).

STRESS PATTERNS AND RHYTHM

We can apply different order-groupings to mutually exclusive sets of properties (e.g. onset, continuation, loudness, as against Lyric) within the same reference-set of sounds. Once fundamental and overall loudness are often organized cooperatively to create stress patterns in sequences. Combined with the organization of durations, we have the group of properties used to define rhythm. Rhythmic structure can be established independently of Lyric structure (as in a dance-hall solo) or parallel to and independent of Lyric order relations. Most pre-twentieth century European art music and e.g. Indian Classical music, segregates out these two groupings of sound properties and orders them in a semi-independent but mutually interactive way (Harmonic Rhythmic in Western Music, melodic/harmonic 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 by slightly stressed patterns, independent of duration organisation, or on a regular duration pattern that is subsequently perceived or seen as varying over time. Stress patterns are a feature over which we can have some control in sound composition. Hence, stress duration patterns at different tempi, or different varying tempi, can be organized in such a way that they synchronize or exactly defined times, these times perhaps governed by larger time-frame stress patterns. (See Chapter 9). (See example 7.19).

By while conscious control of relative loudness or onset characteristics, we can change the perception of the onset states. It might, for example, be possible to create superimposed grouping perceptions on the same streams of amelodic units, e.g. grouped in 5 and in 7 and in 4 and to change their relative perceptual prominence by subtly altering lyric tempos or onset characteristics or in fact to destroy grouping perception by a gradual commercialisation of some inflections. Here field variation is altering order perception. (See example 7.20).
We cannot hope to describe all possible consistant field parameters of all nature-streams because, as the elements of the streams 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 often use a few other suggestions:

1. Variation of sound-type of the constituents (harmonicity-inharmonicity; type of inharmonicity; formatite-type; spinor-charge; 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 nature; 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 temporalisation of groups (moving left, spatial oscillation, spatial randomness), range of temporalisation types, and the time-variation of these.

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

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 increase and decrease 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 large that the density perception itself. Otherwise there is no way to distinguish density from density fluctuation. (See example B.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. 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 entities. 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 relation between entities. Furthermore, slow density fluctuations will tend to be perceived as demand flows — as continuous properties — as compared with towards a simple continuous sound. In these ways we can form bridges between sequences, grain streams and smoothly continuous sounds, establishing auditive links between musically diverse materials. (See example 7.15).

We may also focus the field-properties of a sequence, or part of a sequence by keeping (repeating that portion). Thus a series of pitches may seem unsteady (and may be entirely random across the continuum, establishing no reference frame of pitch.) 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 pitch absence to pitch definitions. (See example 7.16).

The same set-following phenomona applies to the spectral characteristics of sounds. If a meaningful group of weights 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. (See example 7.17).

COMPOSING ORDER PROPERTIES

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

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 organizing pitch-classes which have specific relational properties due to the cyclical nature of the pitch domain (see Chapter 13) it is extended first to the time-domain, which has some similar properties, and then to all aspects of related 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 (claiming that they are preceptually order-related are often open to dispute). All the insights of serial or permucational thinking about materials may be applied to any type of sequence, provided that we ask the vital question, "Is this design usable?" and, in what sense is it usable? — as an explicit in-ordering relation — or as an underlying field concept?

In traditional score music, the manipulation of set-order relations is usually confined to a subset of the properties of the sounds e.g. pitch, fixed sound values, a set of sound 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 monochromatic orchestral ensemble — Weber's Opus 21 Sinfonie provides us with such an instrument-to-instrument line and sets it harmonically. We could, however, permute the instrumental sequence and set the pitches. Formate sequence e.g.

\[\text{bs-ba-b} = \text{ds-ds-d} = \text{bs-ba-b}\]

and instead characteristic sequence e.g.

\[\text{bs-ba-b} = \text{bs-ba-b} = \text{bs-ba-b} \text{ etc.}\]
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 phonetics of a natural language, but traditional musical practice is usually concerned with working on subsets of these norms and exploring the particular properties and possibilities of these subsets.

In this context the size of the field is significant. Musical settings of prose, for example, may treat the Hiphop (and rhetoric) essential to terms of a small overtaken reference set which is consciously represented in subsets (e.g. chord formations over a scale) and remembered (motivic variation), whereas all phonetic material establishes on such small time-frame field and order properties. One tune in sound as referential language, and field and order properties are on the very large timesteps of extensive language utterance. In this situation, the tune is perceived as being in a separate domain to the "speech".

Poetry, however, through assurance, affiliation 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 phonetic and traditional Hiphop and traditional musical concerns and these conventions have been explored by sound poets (Antithphanumeric etc) and composers (Berto etc) alike. As we move towards poetry which is more strongly focused in the nuance of words, or just of syllabic stresses, the importance of small-scale reference-frame and order sequencing may become overriding (e.g. Schottter Grasmus). (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-based, or a traffic recording in a tunnel with a particular strong resonance, or a conversation in Japanese) requires that some field properties of the source sounds will filter into the resulting sequence: a defined Hiphop reference set and a flute spectrum (with or without overt characteristics); in the latter case the field is altered, rising noise-based within a given range, the resonance characteristics of the tunnel, the spectral characteristics of the phonemes of Japanese and perhaps the exact characteristics of the specific voices. These field properties may shift to emphasize the boundaries of the compositional domain (a piano in a piano in a piano) or conversely become part of the substance of it, as we transform the field characteristics (piano -> "bell" -> "gong" -> "cybonet" -> unvoiced sibilant etc.). (Sound example 7.13).

Consequently 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 division with indeterminacy, 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 alone-with processes like reverberation, delay, small-time-dramatic brassage, spectral filtering, spectral shaping, vocalization, or granular reconstruction. = simply through time-continuous (of various sorts) so that the sequence succession = = below the grain-time-frame and the percept becomes continuous. We thus move a sequence rapid density fluctuations. In extreme dense textures, the density fluctuations between = may approach the size of large grains =. A = grain granular. (Sound example R.14).

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

In Sound example R.15 (from Red Noise 1977; a pre-digital work), a working relation was interpreted with (a modified) xylorion noise via this process of temporal thickening. In Sound example R.16 (from Vex 1978), behind the stream of complex vocal multiphonics, we hear the sound of a crowd emerging from a vocal-like noise based which is itself a very dense texture made from that crowd sound.

When this process it to its density extreme we can produce white-noise. When walking on snow in Blizzard conditions it is sometimes possible to become completely disoriented. If the snowfall is sufficiently heavy, all sense of large outlines in the landscape are lost in the walker 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 mere 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-noise.

In Sound example R.17, a dense texture of vocal sounds white-noise and the resulting noise-band is then filtered to give it spectral pitch while it slides up and down. The pitchness is then gradually removed from all hands except one, and the densest document 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 simply overlooked, or even have nested dependency properties. Thus a texture may consist of events whose onset are entirely randomly distributed in time but whose event-elements are clearly rhythmically ordered within themselves, or, conversely, events may begin as the Hiprecipe of a clearly defined Hiphop field while the event-elements have Hiprecipe distributed randomly over the continuum. (Sound example 7.18).
Using average techniques, we may achieve similar results, if the segment size is suitably large and we work on a clearly spectrally-evolving source. Using several segments with similar segment length settings but different ranges (see Appendix 4A-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, reveals unpredictably to already revealed material (Sound example 7.5).

Brassage with pitch variation of segments over a finite Pitch set (possibly cyclically) could establish an Hispach-sequence from a pitch-continuous (or other) source (Sound example 7.6). Even brassage with spatialization (see Appendix 4D) will separate a continuous (or any other) source into a spatial sequence and as long as they are clearly delineated, such spatial sequences can be meaningfully 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 reorientated 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 large time-frame (Sound example 7.8).

Alternatively, copies of the sequence, or groups of its constituents may be arranged in sequences (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 disassembling all but the most persistent spectral properties of the sequence in a way of complexity (the complex in fact becomes a simple sensory percept). (See Appendix 4E). (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 sentences 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 this 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 Pitch 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 characterized by order properties. In existing musical languages, certain sequences of notes will be commonplace, others exceedingly unlikely. In a particular natural language certain groupings of constituents (e.g. "no" 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 patterns where pattern is only hinted at!
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 liedorgan, klagerobahneschiller, passing between the instruments of an ensemble, a "trombone" on a multi-instrument percussion set, etc.). We can also construct disjoint sequences of arbitrary sounds by simply splicing them together (e.g. a dripping tap, a car horn, an older note, a cough - with environmentally appropriate or inappropriate loudness and reverberation relations or not) or by modifying existing natural sequences (time-construction of speech or music, for example). (Sound example 7.3)

Naturally occurring sequences cannot necessarily be accurately reproduced by splicing together (as in the studio) constituent elements. The speech stream in particular has complex transient properties at the interfaces between different phonemes which are not (yet) entirely the subject of intensive investigation by musicians in speech analyses. To synthesize the speech stream it may be more appropriate to model all the transitions between the elements we tend to notice in our writing systems, rather than those elements themselves (this is Diploma synthesis). Starting from the separate elements themselves, to achieve the flowing unity of such natural speech as speech, it may be necessary to "manage" a partly spliced-together sequence. A simple approach might be to add a little sub-tie reverberation. However, for the present discussion, we will ignore this sub-tie 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 & DESTROYING 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 reassembled, do not retain the spectral--continuity of the original. (Sound example 7.4). An existing speech--stream can be constantly modified in the same dynamic manner (depending on where we cut it) the phoneme--continuity. We may do this by chopping up the sequence into component segments and modifying them (e.g. speed halving, seen below), or by selecting segments to cue on (or random as they might overlap other chosen segments) and modifying 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, magnificently or randomly, before constructing the new sequence. Alternatively we may cut out maximal or essential intervals, modify each in a non-progressive manner (different alterations, pitch shifts, etc) and reconstitute the original sequence by replacing the elements together again, but they will now have discontinuously varying imported properties. (Sound example 7.6).

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 temporal--time--frames to dimensions beyond that of the beats and towards--towards the time--scale of a whole piece.

WHERE DOES DENSITY PERCEPTION BEGIN?

In the previous chapter we discussed--stream--streams which were apparently dense but which might retain field properties in other dimensions (like Hptch or formant--eyes). 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 coalesced 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 (net time--time but pitch--wise). In this case we can begin to see that the coalescence 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 Hptic field or Hmonic reference frame, though we may continue to be aware of the range limits of the field. (Sound example 8.1).

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. Hptic organisation).

Composititionally, we can create sequences of events that gradually lose a (transient) sense of rhythm. That we may begin with a computer--quantized rhythmic sequence which has an "intentional" or "mechanical" precision. Adding a very small amount of random scatter to the location--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 pattern is lost. The time--sequence is amorphous. Once this point is reached we perceive the event succession as having a certain density of event entries - 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 process (which is being articulated by random sampling) in a long way into the sequence. We are given a reference frame by the initial event 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 atypical.

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. (Issued example 3.2).

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–quantization grid). Allowing event–onsets to be displaced randomly by very small amounts from this time–reference–frame, we initially retain the aspect 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 pitch reference frame. Here we would begin with events confined to an pitch set (a Harmonic field), then gradually randomize the timing of tones from the pitch set, slowly destroying the pitch 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 relative over such a reference frame. Destroying rhythmicity is hence analogous to destroying the percept of pitches which also retain intermittently 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 semitones for scales played in the Western tempered scale, the onsets 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 in 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 continuous–level perception (or lack of it) into longer duration 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 pitch 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 pitch percept by the randomisation of onset 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.
We can go further than this. We might arrange the grains in some new order (grain reordering). A rising succession of pitched sounds might thus be converted into a (melodic) sequence. (An 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-sequence 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.

MEASURED, COMPARATIVE & TEXTURAL PERCEPTION

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

In the ensuing discussion we will use the term "duration" to mean event-count-sequence-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 fluctuations).

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 compositional time-frame) entire sequences are put in motion (rhythms) while simultaneously establishing a time-frame at the larger metrical unit (e.g. the bar) in which longer-duration orde-sequence 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 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" what 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 measured judgement 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 theme, the employment of tempo and metrical grouping in a particular piece). If we do not have a reference-frame, 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 (e.g. the repeated quaver at a tempo of 120) we can acquire with a fair degree of precision the various multiples (dotted crotchet, minims, etc) and integral divisions (semiquaver, demisemiquaver, triplet semiquaver) of this unit. (See Diagram 1). Hence we can, in many cases, make 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 of Diagram 2 we will probably have a clear-cut sequence of duration values in the perceivable measured ratios 1:2:3:4:5. Our perception is quite 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 perception of "swelling down" but we do not (normally) have a clear perception of the exact proportions appertaining between the successive events. Furthermore, no two performers of this event will perceive the same set of measured proportions between the constraints. This is, then, an example of comparative perception.
DIAGRAM 1

![Diagram 1](image)

DIAGRAM 2

![Diagram 2](image)

DIAGRAM 3

![Diagram 3](image)

DIAGRAM 4

![Diagram 4](image)

**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 time-speed variation or spectral time-contraction (with no grain pitch alteration) may force the grain separation under the maximum time limit for grain-perception and the granularity frequency will eventually emerge as a pitch. (Sound example 4.39). Alternatively, by speeding up the sequence rate of grains without changing the grains to major time-shrinking; see below) we will breach the grain-perception limit. The event will gradually become a continuous flux. In this case a related pitch may or may not emerge. (Sound example 6.13B). Time-reversal 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 blurr the sound 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-shrinking: Sound example 6.15).

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

3. gradually shifting the pitches or spectral quality of different grains differently, so the grain-stream becomes a sequent (granular rendering: Sound example 6.13).

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 pitchshifting 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 limit grains in a similar way to the way we deal with note events. By the appropriate use of gating, cutting and replacing or running, which may be all combined in a single sound processing instrument, we can refine the grains in a sequence using various manoeuvres, techniques (also by a fixed factor, multiplied arithmetically, geometrically or exponentially, relating grain locations about their mean, shrink the time-frame in similar ways and so on: granular time-warps). 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 reversion), i.e. reverse the individual grains themselves (sound reversion), producing what a traditional composer would recognize as a retrograde (as opposed to an accretion; see Chapter 5). Then a grain-stream moving upwards in pitch would become a grain-stream moving downwards in pitch. (Sound example 6.17).
CONSTRUCTING GRAIN STREAMS

Grain-streams appear naturally from intensive sound-production—any kind of roll or trill on drums, key percussion or any sounding material. They are produced vocally by rolled "v" sounds of various sorts in various languages, by lip-fretting and by struck cymbals. Vocalv, such sounds may be used to modulate objects (such rolled "v", whistled rolled "V", flauto-tongued woodwind and brass etc). The rapid opening and closing of a resonating cavity containing a resonant source (e.g. the hand over the mouth as we sing or whistle) can be used to naturally grain-stream any sound. (Sample example 6.6).

In the studio, any continuous sound may be grain-streamed by imposing an appropriate on-off type loudness trajectory (envelope), which itself might be obtained by envelope following another sound (see Chapter 13). (On-off might be anything from a deep irregular fluctuation to an abrupt on-off gating of the signal). A particular delay c is used to achieve the effect to grain a loudness trajectory on a sound tied to the waveforms or wavecubes it contains (wavelet enveopolyeg). If each on-off type trajectory is between 25 and 100 wavecubes in length, we will hear grain-streaming. (Below this limit, we may produce a russian or special "crumming" of the source sound). This effect uses the grain-streaming in the internal proportion of the source-sound as, for example, the grain-streaming manifest itself if the perceived pitch falls. This suggests to the ear that the fallc portamento and the instrument are causally linked and intrinsic to the source-sound, rather than a compositional affectation. (Sample example 6.4).

Alternatively grain-streams may be combined by placing together individual grains (1). 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 uncorrelated grains over a small range (serration) produces a more convincingly natural result. (Sample example 6.9).

More compositionally flexible, but more perspicuity, is to use a mixing program so that individual grains can be panned and ordered, then repositioned, replaced or modified using mere-instruments which allow us to manipulate mixing instructions (maleflocking) or to generate and manipulate time-sequence (sequence generation).

In this way grain-streams can be given greater acrobatics or rambunctious of different shapes and be slightly randomised to avoid a mechanical result. (Sample example 6.6). Similarly, the grains themselves can be sequentially modified using some kind of envelope (see Chapter 4). or spectral transformation tools (e.g. destructive transion through wavecutter distortion, wavecutter harmonic distortion or wavecutter-averaging; see Chapter 3) combined perhaps with rebuchening (see Chapter 12) to generate a set of immediate sound-streams. (Sample example 6.7).

Short continuous sounds can be extended into longer grain-streamed sound by using bounce with appropriate grain (time-frame duration) segment-ize. (Sample example 6.8). The granularity of the remaining source can be exaggerated by correcising (see Chapter 10) and the regularity of the result mitigated by some of the grain-stream manipulation tools to be described below.

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

A more interesting example is generated by the sequence in Diagram 4. If this rhythm occurs in a context where there is a clear underlying susiquence reference-frame, and the sequence is played "precisely" as written, we will perceive the 3:1:2:3:1:2 etc, sequence of duration proportions clearly—our perception will be measured. However, in many cases, this time pattern is encountering where the main reference frame is the crooked, 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 alteration of short and long durcations which, however, are not necessarily perceived in some measurable proportions. The score may give an illogical setrion to a 5:1 definition, but we are concerned here with the percept. (Sample example 9.4).

The way in which such crooked beams are divided is one aspect of a sense of "swinging" in certain styles of music. A particular drummer, for example, may have an almost completely regular long-short division of the crooked in the proportion 37,26. We may be aware of the regularity of taullc beat and appreciative of the particular quality of this division in the sense of the particular sense of swing it imparts to toaller playing while remaining completely unaware of the exact numerical proportions involved. Here then we have a comparative perception with fundamental qualitative counterpart. (Sample 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 remote to the extent that the precise morphology of each viscal note cannot be specified in the notation and varies arbitrarily over a small range of possibilities, as does the vibration of opera singers. In sound composition, these factors can be precisely specified and, if desired, random to the level of comparative, or measured, perception.

More importantly, our rhythmic example using dorted rhythms illustrates the second aspect of our discussion. For in this example, perception at the level of the crooked remaine measured—the music is"time". However, simultaneously, in a smaller time-frame, our perception has become comparative. We can use the same division into time-frumes if we look again at the idea of "independability factor" propoed by Clarence Barlow (see Chapter 7). As discussed previously, we can deflir, over a reference frame of quarres, the relative independability of each note in a 3/4 or a 3/4 pattern. Limiting the probability of occurrence of a note to its independability factor allows us to generate a strong 0.9 grouping for, or a strong 0.25 grouping, or an ambiguous partc between the two. Once every quarre becomes equally probable, however, the sense of grouping breaks down altogether and at the level of 4-groups (1-groupings, or any grouping) measured perception is lost. Our perception returns to the field characteristics. In this case, however, the field is defined by a smaller set of durcations, the quarers themselves. So at the smaller time-frame, we remain 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 measurable 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 limits of grain perception but comparative judgement of grain—durcations, especially in more demanding proportions (1:5 to opposed to 1:2), seems to break down well above this limit.
CHAPTER 6

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 were cyclically), rather than as a sequence of changing but definite sound events. Once, however, the individual spectral events 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 rate speed-of-normal-speed 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 dented by the onset distortion but the remainder of the sound may appeal spectrally satisfactory. If we time-stretch a discontinuous sound, however, we will be drenched everywhere by onset distortion as the sound is a sequence of events. Often we want the sound (e.g. in the real environment, a drum roll, a speech-sound) to be delivered more slowly to us without the individual attacks (the drum rolls, the onsets of consonants) being accentuated and hence transformed. We work on the inference 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 spectral timbre of the onset events is not altered. 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 merely a simple grain rapidly repeated. Even where this (next) ideal limit is approached in naturally occurring sounds (e.g. low confluence streams) 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 grains (speech, 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 compared to have) overall continuation properties (disjunctive, supplementary and found 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.

55
This problem becomes particularly important when different divisions of a pulse are superimposed. To
given an example, if we compare two duration streams in the proportion 2:3 (see Diagram 8) we have a
clear mutual time-frame pulse at the larger time-frame of the crochet. (Sound example 9.6).

If we now re-group 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. 1/1000 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-stumps can be compared in
the two streams) that we have 2 streams of similar but different event density.

With 3 superimposed stems 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 crochet best into (say) 3, 11, & 13 in (crochet = 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
materialized in practice?

No human performance is strictly rhythmically regular (see discussion of quantized rhythm above).
When three streams are laid together the deviations from regularity will not flow in parallel (they will
not synchronize like the parallel micro-anticipations of the parts 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 manner of how much an event is displaced from its
"one" position. A factor of 1 means it is displaced by a whole unit. In larger time-frame norms, a
quaver in a sequence of quavers would be inaccurately placed by up to a whole quaver’s length. By
someone’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 unit 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).
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 8.23). We may also cycle round a set of pitches in a very small band providing a subtle "stuttered vibrato" inside the continuum (particularly if grain-size is slightly random—avoid to rhythmic common in perception). (Sound example 8.24). The pitch band can also be progressively broadened. (Sound example 8.25). The loudness of segments can also be varied in a random, progressive or cyclical way. We might also specialize, or progressively specialize, the segment components, moving from a point source to a spread source. (Sound example 8.26).

Eventually, such evolved manipulations (and their combination) form a continuous or coherently segmented source to discontinue into a mass of acoustic events. The ultimate process of this type is in Sound templating which completely deconstructs the original sound and will be discussed in Chapter 7.

**CONSTRUCTED CONTINUATION: GRANULAR RECONSTRUCTION**

We may generalize 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 replacing these tail-to-tail, they are used as the elements of a growing texture in which we can control the density and the time—randomisation of the elements. (See Appendix p73). (Sound example 8.27).

In this way we can overlay segments in the resulting stream, or, if segments are very short, introduce monaural silences between the grains. This process, especially where used with very tiny grains, is also known as granular synthesis (the short-samples between sound processing and synthesis are fluid) and we may expect the spectral properties and the time—characteristics of the grains to influence the quality of the resulting sound stream, alongside imposed stream—randomisation (density, pitch spread, loudness spread, spatial spread etc.). This process then passes over into texture sound, and it discussed more fully in Chapter 8. (Sound example 8.28).

With granular reconstruction, if we keep the rate and pitch—spread small, we may expect to generate a time—stretched grain sound which is spectrally thickened (and hence often drawn easily, at least less focused), the degree of thickening being controlled by our choice of both density and pitch bandwidth. But as rate, density and bandwidth are increased and segments duration varied, perhaps progressively, the nature of the source will come to have a less significant influence on the grain sound. It will become part of the overall field—properties of a texture stream (see Chapter 8). (Sound example 8.29).

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 overlap (apart from the length of the edit themselves) when they are reassembled. In practice, normal brassage/harmoniser processors are a certain degree of overlap of the adjacent segments to ensure sonic continuity and avoid granular artefacts (when these are not required) in the resulting sound.
samples 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 p145). (Sound example 5.16).

Zigzagging, however, can move from any point to any other within the sound, constant 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 repetitions. By altering the zigzag points (notably from rig to big), we may vary the length (and hence duration) over the repeated sequence. In this way zigzagging can be used to generate non-mechanical stochastic properties, or (in a longer timescale) dispersed-acumulation (see above) effects, within a sound-continuation. (Sound example 5.17).

Sounds can also be artificially continued by brevage. In the brevage process successive segments are cut from a sound and then rejoined together to produce a new sound. Clearly, if the segments are selected 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 p46-B). (see note at end of chapter).

As discussed previously, grain time-frame segments will produce a simple time-meshing of the sound (harmonic effect). Slightly larger segments may introduce a granular percept into the sound as the perceived evolving shape 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 where operating on a segmented source (a melody, a speech stream) will result in a systematic collaging 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 brevage can be generalized. Using non-rigid 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-down) or articulatory (short but longer time-frame) aura to the time-meshed 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 p46-C). In this way, echo percepts are randomized further. Softly controlling this and the previous factors, we can create rich fields of possibilities from the small features of sounds with evolving spectra (especially sequences, which present us with constant and perceptibly evolving spectral) (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-B), Now, as we proceed, all the previous properties of the sound become grist to the mill of continuation-production. In the case of e.g. a long, melodic phrase which we brevage 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 Echotic possibilities in the goal sound. (Sound example 5.21). On a smaller time-frame, the qualities of a highly characteristic event can be underdubbed over the entire ensuing continuation. (Sound example 5.22).
In our second example, the maternal 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 more Maggie.

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-to-the-nine-of-1 grouping, and the 6th unit in an 11-to-the-nine-of-one grouping, over the same crochets, at [tune] = 120 are only 0.5436 of a crochet, or 0.1268 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 can see, 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 sense specifying a density flow with the limitation 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 mention 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 perceptions 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 "arbitrarily" related groupings, even given the intrinsic "imaccuracy" 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-processing in generating the sound sequence, we may be able to perceive a very precise "density flux" in the combination of these versions, at least if they are repeated a sufficient number of times. (Sample example 9.5).

Our comments about notated music are further reinforced if we now organise the material internally so as to contradict any eventual reinforcement at larger pole (e.g. bar) interface. 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 regularly in a yet larger-time-frame, e.g. the 4-bar frame. (Diagram 14).

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

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

process also can change a rapid continuation into something in a phrase time-frame, for example, formant-gliding comments ("v", "v" in English) become slow formant transitions ("v" --- "v" and "v" --- "v") Conversely, a construction structure can be locked into the indivisible character of a grain, through extreme time-contraction. (Sample 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 nonmeasurably unstable pitch and spectral properties in a short event. It may also be combined with filtering to sustain specific pitches or spectral peaks. It provides a new dispensation continuation for the sound. (Sample 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: hallways or swimming pool enclosures. Reverberation in fact results from various delays (due to travelling in indirect paths to the ear, by bouncing off walls or other surfaces) and specially delayed (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 mimicker has the added advantage that we can specify the dimensions of unlikely or impossible spaces (e.g. infinite reverberation from an "ininitely long hall", the inside of a broom in use, an orchestra etc.) There is then 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 auditory or formal continuation properties of a sound. On the contrary it will alter these. Moving pitch will be extended as a pitch band, loudness variations will simply be averaged in the new dispersion. We can, of course, post-tune, and addulatory or form continuation properties (vibrato, tremolo, envelope) to a reverberation extended sound. These will at first help to unify the perceptual initiation-reverberation in a single sound-event. The reverberation part of the sound will appear to be part of the sound production itself, rather than a aspect of a sound box or characteristic of a room. (Sample example 5.15).

Inverter—reverberation models have in fact been used recently to build synthesis models of acoustic instruments from vibraphones, where the separation might seem natural, in Yahava.

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 reverse-read starting at the point where the preceding read ended. The overall 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 points in the middle of the sound until told to proceed to the end. This "alternate looping" technique was used in early commercial
crowd) by applying an appropriate loudness trajectory (enveloping) with perhaps more subtle parallel delays (specific enhancement by flaring, harmonically shifting by spectral smearing or subtle delay). Moreover, these visually created shapes can transcend the boundaries of the physically likely. Sounds which are necessarily lower in the real world (reverberant whispering) can be usefully local while sounds which are associated with great dimensions - e.g. the crashing together of large, heavy objects, the foresting of metal surfaces, can be given a plenitude of activity.

Moreover, we can extract the properties of existing patterns 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), mapping it to the original sound (envelope manipulation), or transforming it to other sounds (enveloping or envelope substitution; see Appendices p99).

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 unrecognizable phenomena (the dripping of a tap) or working in stereo, the character of a whole field of naturally occurring events (e.g. traffic flow, the swirling of bees etc.). The extracted general information can then be modified and mapped to the same material (envelope transformation followed by envelope substitution), or applied to some entirely different musical phenomena (envelope or envelope substitution; see Appendices p99) or stored for use at some later date. All such manipulation of the loudness trajectory are discussed more fully in Chapter 10.

Sounds may also have a specific spatial continuation. A whole chapter of the Sonic Art is devoted to the exploration of spatial possibilities in sound composition. Here we will only note that spatial movements can be a type of formal continuation applied to a sound, that it will work more effectively with some sounds than with others and that it can be transmitted from one sound to another provided these phenomena are borne in mind. Many sounds, sound-streams and sound 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 indistinguishable quality of grain or of sound source. (Sound example S12).

The movement from mono into stereo may also play a significant role in dynamic incorporation (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 result in continuation properties intrinsic to the sound (reverberation, time-stretching, some kinds of brassing) while others impose their own continuation properties (lipping, granular-ambience and other types of brassing).

The most obvious way to create continuation is through time-stretching. There are several ways to do this (brassing/granularization, spectral stretching, wavelet 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 indivisibility qualities of a grain into a perceptible time-varying structure - a continuation. In this way an indivisible entity can be made to reveal a surprising morphology. This
LARGE-DURATION REFERENCE FRAMES — THE NOT SO GOLDEN SECTION

At the other extreme, music can be constructed as a set of **small** time-frames in which at one level (e.g. a minim or quaver time-frame) order-sequencing is set in motion (motions) while at the same time establishing a time-frame (e.g. mordent form) in which longer duration order sequences can be set up.

Within the time-frames fast-dominant/quaver to 1- or 2-minute, hierarchical systems of intersected 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 tenacious limit to perceived perception. Though many composers writing musical scenes would make their representations on the idea that was clear and reserved to precise proportions in larger time-frames (especially those derived from the Fibonacci series and Isomorphology (Golden Section) there is little evidence to support this. Such proportions certainly look once to ones and provide consistent ways to subdivide larger time-scales.

The exact proportions offered by the Fibonacci series have the particular advantage of being notate at least 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 such proportions in a measured sense, or do we perceive them in a comparative sense (1 to 1/2/3/4/2/5) while the exact measures make the task of laying out the zero possible. If we consider the golden section itself (the limit of the Fibonacci series) this will come 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 rules, 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 in using space-segments to such irrational proportions and hence, hypothetically, a perfect Golden Section existing might be achieved in a studio composition. Whether we are measurably pursuing this handwork is quite a different question.

On the relatively small time-frame of “swinging” (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 can use the proportion of both longer term time memory and the influence of smaller scale events (part A: our sense of the passage of time (walking the minutes for this train seemed an interminable time ... reading the newspaper, an hour gone by without any 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 order 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 comparatively perceive these proportions (e.g. 3/4 to 1/3) or merely that we can comparatively perceive them all at lying within the range 1 to 1/2/3/2/3/5. More dramatically, because the Golden Section is an irrational quantity, then, by ideally we should separate format data before adding some new vibrato property to a sound, reanalyzing the separate motion of the formants on the pitch-aided sound. (See former preserving spectral manipulation: Appendix p27). To practice we can often get away with more tape = speed variation transposition of the original sound source.

We may extract the ongoing evolution of underlying properties using pitch-tracting (Appendix pp17-21) to envelope following (Appendix p58) and apply the extracted data to other events. We may also modify (e.g. expanding compression: Appendix p69) the underlying properties of the sound-source. Alternatively, we may define modifications in pitch (vibrato) or loudness (sustain) in a sound, with control over time-varying speed and depth. In this way we may, e.g., improve voice-like articulations on sounds with static spatters, or non-vocal environmental sounds (e.g. the sound of a power drill). (Sound example 5.7)

We may also produce extreme, or non-naturally—occurring, articulations using extremely wide (several octaves) vibrato, or pitch tremolos to the point of sound granulation. At all stages the overall resonances (granular variation of sound and depth) will be important compositional parameters. (Sound example 5.8)

In the limit (very wide and slow) tremolos becomes large-scale loudness trajectories and we may progress from unfolding continuous to fast oscillations (true voice vio). Similarly, vibrato (very wide and slow) becomes the forced oscillations of moving pitch. We thus move from the undulatory oscillations of a static pitch to the rhythmic pitch motion structures in which pitch has no further significance. (Sound example 5.9)

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

We may also imagine undulatory oscillation of a sound’s envelope, reinterpreting it in its harmonic/subharmonic dimension, its sub-structure to subharmonic dimension, or in a formant-position dimension (spectral undulation). Formant undulations (like hard-python, head-shake flutters or “pitchy” articulations) can be produced and controlled entirely vocally. (Sound example 5.11)

FORCED CONTINUATION AND ITS DEVELOPMENT

In any system where the energizing source has to be continually applied to sustain the sound (brushed using, blown nose, speech) the activating energizes continuous current over the evolution of the sound. With an instrument, a player can force a particular type of continuation, a crescendo, a vibrato or tremolo, or an overblowing, usually in a time-constraining fashion. In general, the way in which the sound changes in loudness and spectrum (with uncurved or bowed sounds etc) or sometimes in pitch (with more generalized as in a town 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 then be thought of as physical gestures translated into sound.

Any gestural shape which can be applied by breath control or loud pressure on a wind or bowed using instrument, can be reproduced by any arbitrary sound (a sustained piano tone, the sound of a dense
definition, there is an smaller time-frame over which it can be reasonably perceived. This is part of the
definition of a mathematically irrational quantity. Hence, by definition, we cannot reasonably perceive a
Golden Section, though again we may comparatively perceive that it lies within the bounds of 1 to
\[ \frac{\sqrt{5} - 1}{2} \approx 0.61803 \text{ or } 0.61805 \]

I am happy to accept that Fibonacci derived ratios give a satisfactory approximate perception of 1 to
1/1.61803-2/3 (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 range. The longer the time-frame
the larger this range of irrationals becomes.

I am certain of two things. Firstly that when psycho-acousticians are detailed tasks, even on
composers who regularly use these devices, they will discover that these illusions 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 icon of a modern musical mysticism.

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-placements,
exact pitch-values, exact spectral structure 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 made in context. Is the musical
organisation focusing our attention on these exact timing aspects or is the overall sound important, 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 not central to the way in which
we scan the musical events for a sense of mutual connections and contrast?

It is always possible to claim that everything is important. In particular, composers are especially
good at claiming that anything is an important constituent of what the listener hears. While it was an
important part of the work 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 waveshape has at least one sample
of value 327. This, by itself, will provide no perceived coherence within the sound world I create. To
show that it does is simply dishonest. The composer must then be able to discern what is and what is
not perceptually 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, tuned to 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.
The structure of dissonance and accumulation can be combined to generate more interesting composition structures. By applying copies of one segment of a dissonant sound in a back-and-forth fashion so that the reverbed version exactly dovetails into its forward version, we can create an ebh and flow of spectral energy (see Diagram 1). By time-variably time-shifting these portions, we can avoid a merely time-cyclic ebh and flow. More significantly, the closer we cut to the onset of the original sound, the more spectral movement our accumulation will gather before eliminating into the dissonant 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 reminiscence law was spread from the Japanese governent habit of eating a fish which is poisonous if the fish is not close to the liver in status. Does efficiency, out of brevity, ask for the fish to be cut within half the length of the piece, someone with Titus-Minnesingers. Sound composition is, fortunately, a little less dangerous. (Sound example 5.3).

Again, time-varying, spatial situation or different types of spectral alienation (see Chapter 3) may 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, modulation of pitch (vibrato) and undulation of loudness (breath). Undulating continuation is related to physical activities like shaking and it is no accident that a wide, still-like, vibrato is known as a "shaker". These variations in vocal sounds involve, in some sense, the physical shaking of the diaphragm, larynx or throat or (in extended vocal techniques) the rib cage, the head or the whole body (1). 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 occurring vibrato and tremolando, there is moment-to-moment intensity of undulation speed (frequency) and undulation depth (pitch excursion for vibrato, loudness fluctuation for tremolando) which at no instant is obvious until we create artificial vibrato or tremolando in which those 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 am music cultures, subtle control of vibrato speed and depth is an important aspect of performance. Even in Western popular music, glissando upwards on an Ohojik as vibrato is added, it is a common phenomenon. (Sound example 5.6).

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

Sounds in the met world which do not simply die away to nothing require some continuous energy input such as bowing, blowing, singing 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 fluctuations causally of the sound.

We may create the illusion of energy flow with a continuous, growl-sustain, sequential or layered sound by simply imposing a loudness trajectory, or envelope, on it (envelopeing) (See example 10.1). Such a trajectory may be derived directly from the activity of another sound (preserved or a live performance) (envelope following). Enveloping is particularly useful for creating a loudness anacresis (a crescendoing onto a key tonic event), a particularly musical device as most natural sounds have precisely the opposite loudness evolution (See example 10.2).

Conversely, we may begin with a sustained sound (continuum, growl-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 (for otherwise not in motion) and then left to resonate, especially if the onset of the sound is emphasized. With simple sustained sounds this can easily alter the apparent physicality and character of the source (see below). With sounds which more easily 'recall their origin' (e.g. speech) we produce an interesting dual percept. (See 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 transform the trajectory to an entirely different sound (enveloping or envelope substitution). Sequences of events which are externally strongly characterized by loudness articulation may thus be "reconstituted" with an entirely different sonic identity (See example 10.4).

We may also modify the loudness trajectory we have extracted (envelope transformation) and rapidly Play the original sound. Note that we must do this by envelope substitution and not simply by enveloping (see Appendix E parts 4 & 6). 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 compression), extending or contracting the associated gesture (see for example the discussion of time-warping 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 correted (see below).

WHAT IS CONTINUATION?

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

Some types of sound-continuation are, however, quite special. Sounds made from a sequence of perceived rapid counts (grain-streams), sounds made from sequences of short and different elements (sequences) and sounds which dynamically transform one set of characteristics into a quite different set (dynamic interpolations) all have special continuation properties which we shall 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-sequences, sequences and dynamically interleaving sounds. They are not mutually exclusive.

DISPERSEHÓN CONTINUATION & ITS DEVELOPMENT

Certain natural sounds are initiated by a single or brief pause (picking, short blow or rh) and then continue to evolve because the physical material involved has some natural internal resonance, (undamped natural strings, bells) or the sound is maintained by a continuous resonance (alum drum, marathon bell scenarios). As the medium is no longer being excited, however, the sound will gradually become quieter (not inevitably, for example the sound of the tumb-tum may grow louder from 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 persist for longer than any other pitch component not having such a resonating mode. We will call this mode of continuation ambient dispersal (See 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 special characteristics and ending abruptly at a point of (usually) maximum loudness. The only real-world comparable experience might be that of a broad-source approaching us at a great distance and suddenly stopping on reaching our location. We will call this type of continuation as accumulative.

Accumulations are more striking if they are made from non-linear dispersals. The decay ofloudness and spectral energy of a piano note from the close to linear, so the associated accumulation is little more than a crescendo. Chords or tumb-tums to other complex-spectra events (e.g. the resonance of the unstruck piano frame when struck by a heavy mallet which have a much more complex dispersal in which many of the initial high frequency components die away rapidly. The associated accumulation begins very gradually but sometimes in spaced "moments" linearly, the sound, generating a growing sense of anticipation. (See example 5.2).
FINE-GRAINED ENVELOPE FOLLOWING

The result of the process of envelope following will depend to a great extent on how the loudness trajectory is analyzed. 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 6.5(b)). 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 amplitudes 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 coarse 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 fine monitory silences between the grains (conversation). 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 gran time-frame has little meaning. The instantaneous value (pressure) of the sound wave changes from positive (above the mean) to negative (below the mean) 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 on one (or only a few) wavecycles, or wavefronts, 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 6.5(b)). But if we gradually expand the number of wavecycles or wavefronts we cause to fall under our loudness-changing envelope, we pass from the spectral domain to the time-domain of grain-structure and eventually to that of independent events. These transitions are neatly illustrated by waveenveloping. In sound example 10.6, the number of wavefronts 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, and 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 highlight 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 18/7).

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 direct. Combining this with triggered spatial position control we might achieve a whole stereo-versus-reverberation-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. As an injector, any sound above a certain threshold loudness is reduced in level to that threshold value. This means 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 it 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 parental trajectory of a sound, thereby e.g. "calming" a hyperactive sequence of events.

As expander, it's 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 p67.)

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 added 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 proximities characteristics (close-related, or mixed at a distance such that room ambiance is significantly integrated into the sound).

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 (onsets, 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 instances in music will eventually destroy onset characteristics, whereas stereo separation will allow the ear to discriminate event onsets even in very dense textures. (Sound example 4/9)

ALTEDER 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 remain in "voices" when continuous information contradicts our initial intuitive assumption. The piece Vox-3 uses this causality transfer throughout in a very convincing way but it can operate on a more immediate level.

Listen first to Sound example 4/12. A vocally based event transitions 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-crossing)in this sound where the vocal initiation is shifted onto its non-vocal (that voice derived) continuation. (In Vox-3 the vocal-onset transitions are achieved by smooth spectral interpolation, rather than abrupt splicing – see Appendix, U2). When this abrupt change is pointed out to us, we begin to notice it as a rather obvious discrepancy, the "vocal chain" is broken, but in the wider context of a musical piece, using many such vioce initiated events, we may not to easily handle in on the discorodinacy. A more radical causality shift can be produced by onset fusion. When we hear two sounds at the same time certain properties of the neural onsets allow us to differentiate them. Even when we hear two violins playing in unison, we are aware that we are hearing two viioins and not a single instrument producing the same sound stream. At least two important factors in our perception permit us to discriminate 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 combination we can readily separate the sources. Secondly, the onset of the two events will be slightly out of synchronization so that however accurately they are played. Then we can readily separate the two sources in the onset moment.

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

In Sound example 4/11 from Vox-3 this process is applied to three vocal sources. Listen carefully to the first sound in the sequence. The perception is of "two" but also voices, even though the sources are only unperformed voices. This initial sound initiates a sequence of similar sounds, but as the sequence proceeds the vocal sources are also gradually aurally imprinted (See Chapter 3) becoming more and more bell-like in the process.
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 imagine a sudden-onset (snack-food) onset on any sound purely by providing an appropriate onset loudness trajectory, or we can make a sound gradual-onset by giving it an onset loudness trajectory which rises more slowly. In the case time-frame we can proceed from the "snack aesthetic" object to the "snack maximising object" to the "snack maximised object" and beyond that to the situation where the sound appears to rise out of nowhere like the "tinging" of bowl gruel or the like. (Sound example 4.54).

The where the sound has a "snack" quality, we may not just be nit picking but the physical quality of the striking medium. Harder striking agrees tend to excite higher frequencies in the specimen of the vibrated material (compare pitched stick, rather than struck sticks and wooden sticks on a vibraphone). In a generalized notion of physical "hardness" we can move to the onset of any sound. It is an exaggerating the high frequency components in the onset moments we create a more "harsh" or "brittle" attack. (I make no apologies for using these qualitative or analogical terms. Gran :- time-frame events have an indivisible qualitative unity in some respects. We can plot 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 is to mix between 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 upward transposition and the decay is clearly a matter of aesthetic intent and judgment. (attenuation). The transpositions might be in the time frame of the original sound, or time-contrasted (as with age-speed variation; see Chapter 11). The latter will add a 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 correlates of such a process being less clear. This latter fact is not necessarily important as it is 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 (indirectly) in its continuation in the sound of water, or human speech, or a non-epiphanic spectral suggesting physical softness and elasticity. We are not constrained by the photographically real, but our perception is guided by physical sensations even when listening to sound made in the entirely controlled space of sound composition. (Sound example 4.54).

Another procedure is to add noise to the sound onset but allow it to die away very rapidly. We may cause the moment to " resolve" onto an actual wavelength if the modified sound by providing that wavelength with repetitions of itself which are incrementally modified i.e. noisier. This produces a flapped-swing-like attack (sound plugboard) and relates to a well known synthesis instrument for producing plucked string sounds called the Korga Synth Strougl speaker. (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-averaging 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, moreover, we may combine such features in ways which conflict our experience of natural acoustic environments. Thus predominantly recorded sound may be played back very quietly, whilst distant quiet sounds may be perceived very loudly. These features of the sound landscape are discussed more fully in Ch: Source Art.

In sound composition in general, the loudness balance between diverse sources is described in a very general way, mostly either by creating a sound design or by creating a rich texture of a mixture, which is a mixture of active elements. 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 sound substance intermediate between the constituents (e.g. subsonic, see Chapter 12). (Sound example 10.8).

Mixing may also be used to consciously mark 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 eliminating the loudness level of other sources. In popular music the device of doubling 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 "back up" as level as to not 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 intrude a sound of constant energy (a large crowd talking amongst itself, heavy traffic) by inverting the inverse loudness trajectory of the sound on the traffic noise (Sound example 10.10). This procedure will cause 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 onomatopoeic interest of the sound sources themselves!).

PROXIMITY

Loudness information in the real world usually provides us with information about the proximity of a sound source. The sound heard more quietly is usually further away. It is important to understand that other factors figure in loudness 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 perception (presumably the presence or absence of barriers (walls, doors to rooms or entrances to other contains) and the nature of reflective surfaces (hard stone, soft furnishing) contribute to a sense of ambient reverberation. Even air temperature is important (in 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 clearly recorded sounds appear distant, especially when contrasted with the original source. But we may also produce self-contradictory images e.g. clearly implied whispering, which we associate with close-to-the-ear intimacy and hence "quiescence", 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 sounds 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 vowel 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 centimeters) in contrast to the typical range of the visual zoom. (Sound example 10.11).

PHYSICALITY AND CAUSALITY

As we have already suggested, loudness trajectory pays an important part in our attribution of both the physicality of a source (rigid, soft, loose aggregate etc.) and the causality of the excitation (stomp, struck 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 stick-like attack by providing a sudden onset and perhaps an exponential decay. Conversely, a strongly percussive sound may be softened by very carefully "slowing 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.
With acoustic instruments the initiating transition from "off" to "on" is most often a complex event in its own right, a clang, a harsh release, or whatever, with the dimensions of a grain (See Chapter 3) and with its own intrinsic sonic properties. These properties are, in fact, so important that we can destroy the recognizability of instrumental sounds (harp, trumpet, violin) fairly easily by removing their onset portion. It is of course not only the onset which is involved in onset recognition. Sound sources with internal resonance and natural decay (snare drum strings, struck bells) are also partly recognizable through this decay process and it is artificially prevented from occurring, our perception 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 (glockenspiel 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 motions, spectral type (harmonicity, inharmonicity, resonant overtone structure) 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 "snack" as the trajectory implies that all the energy is imparted in an initial shock and then dies away naturally. We can cause the percept "snack object" by imparting such a brief loudness trajectory on almost any spectral structure. For example, a noise-masked 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 masked sound may appear like snack wood. (Sound example 4.1).

If these grains are then spectrally altered (e.g. the various destructive distortion instruments discussed in Chapter 3) we may alter the perceived nature of the "maimant" being "snack". In particular, the more noisy the spectrum, the more "thrum-like" or "crystal-like" but we are retaining the percept "snack" because of the persisting form of the loudness trajectory. (Sound example 4.2).

If, however, we provide a different loudness trajectory (by equalizing) like type 1b, which has a quiet onset and peaks near the end, the energy in the sound tends to grow, which we might intuively associate with rubbing or stroking or some other manner of causing an object into its natural vibrating mode. At the very least the percept is "gradually 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 loud scraping together of aspersions where the evolution of the process is controlled by forces external to the natural vibrating properties of the material e.g. sawing or boring, particularly where those produce forced vibrations rather than natural resonant frequencies. (Sound example 4.4).

Thus with such very brief sound, transformations between grain different are perceived can be effected by the simple device of aligning the loudness trajectory. Combining this with the control of spectral content and spectral change given us a very powerful purchase on the sound-composition of grain-scale sounds.
CHAPTER 11

TIME STRETCHING & TIME-FRAMES

Time-stretching and time-dithering warrant a separate chapter in this book because they are procedures which may breach the boundaries between perceptual time-frames. The importance of time-frames in our perception of sounds is detailed in the section "Time-frames: samples, wavecycles, grams & measurements" in Chapter 1.

Furthermore, the degree of time-stretching of a sound may itself vary from 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, sampling techniques, wavelet repetition (nearest time-stretching), frequency domain time-stretching (spectral time-stretching), and grain manipulation (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 textured streams.

"TAPE-SPEED" VARIATION

In the classical tape-recording 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 invoicing the result as the standard sampling rate). This approach is used in sampling keyboards (1993) and tape-reading instruments e.g. in Classical.

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 p87). This approach is used in the time-domain signal. Similarly, it changes the frequency of the partials and hence also shifts the spectral contour, and hence the timbre. And it time-stretches the onset characteristics, probably radically changing the sound percept in another way (Appendix 111). Although time-warping, pitch-warping and frequency-warping are therefore not independent, this approach has musical applications. In particular, multi-octave (multi) upward transpositions can be used as short time-frame realignments of a sound's onset characteristics (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 more sensitive in this range). The internal qualities of a complex sound may thus be magnified in two complementary domains. Digital recording and transposition is

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 tentative 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. Both need to be struck, dusts blown etc. Some resonating systems can, with practice, be put into resonance with almost no discontinuity (the voice, bowed strings, plucked). Others either require a transient onset event, or can be initiated with such an event (flute or brass reengaging). Other systems have internal resonance - once set in motion we do not have to contain applying 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, CB). Other systems produce internally short sounds as they have no internal resonance. Such sources can produce either individual short sounds (strings, xylophones, many vocal commands) or be activated iteratively (drum roll, rolled "Y", low conduction notes)

Iterative sounds are a special case in which perceptual considerations enter into our judgement. Low and high conduction notes are both produced by the discontinuous movement of the reed. However, in the lower notes we hear not these 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 intense, 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 6.
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. (Sample example 11.2)

Tape-speed variations 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 highest range (using e.g. tape acceleration). As the sound rises, internal detail is lost. With sufficient acceleration, almost any sound can be compressed into a unison-like rising pitch-portamu. This is an elementary way to perceptually link the most diverse sound materials, e.g. they occur in long enough streams for such acclerations to be possible, or a way of creating musical continuity between a complex stream of diverse events and trion-splines focused on pitch pertains. (Sample example 11.3)

BRASSAGE TECHNIQUES

As discussed previously, Brassage involves cating a sound into successive, and possibly overlapping, segments and then reassembling these by repositioning them (in 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 perceivable pitch and spectral properties but no pitch or spectral evolution over time) then the result sound will appear time-stretched relative to the source.

Good algorithms for doing this are currently (1994) embodied in commercially available hardware devices (often known as Brasversors) 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 artifacts of the process. These may, however, be useful as sound-transformation techniques. (Sample 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-stretches the perceptual connection between source sound and goal sound may be remote and may require the perception of modulating 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. (Sample example 11.5)
Braunage 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 then move gradually out of the field of time-stretching into that of generalized braunage and granular reconstruction. (These possibilities are discussed in more detail in Chapter 5 in the section “Contracted Continuation” and in Appendices p54–55 and Appendix p32). We need add only that using Granular Reconstruction with time-varying parameters (average segment length, length spread, stretch 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 more time-stretched and evolves as a texture stream development of that source.

A more time-stretching satisfactory application of the Braunage process to sequences such as speech streams can be achieved by source-segment-synchronous braunage. In this case we need to apply a sophisticated combination of envelope following, and pitch-synchronous spectral analysis to tease the individual segments of the source stream. This can then be individually braunage-time-stretched and respliced together in one operation. Because none of the goal segments crosswise between source segments, we avoid artifacts caused by source segment boundaries in simple braunage (see Diagram 1).

WAVESET TIME-STRETCHING

Time-stretching can be achieved by searching for zero-crossing pairs and repeating the waveforms 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 waveform 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 ±2 time-stretch, by an octave downwards, as in tape-speed variation). For a ±2 (or even ±3) time-stretch, artifacts can often be reduced by repeating pairs (or larger groups) of waveforms (a special case of pitch-synchronous braunage). These kinds of artifacts 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 11.4).

As the number of repetitions increases, other artifacts begin to appear. At ±3 there is often a “rubbing”-like distortion of the sound. With time-evolving and noisy signals, a ±16 time-stretch a rigid stretch of pitched beats is produced as each waveform 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 listen is heard "indefinitely" within the source even at ±4 time-stretch. (Appendix p55). (Sound example 31.7).

It is possible also to interpolate between wavelet durations and between wavelet shapes through the sequence of repetitions. In ±64 time-stretching with such interpolation the new signal clearly glides around in pitch as each “head” pitch glides into the next. At ±64 time-stretch, we are aware more of the “fluidity” of a sound-stream rather than of a continually pantomating line. Even at ±64 time-stretch, this fluidity quality contributes to the percept in an intangible way. (Sound example 11.8).

CONCLUSION

In a sense, almost any manipulation of a signal will alter its spectrum. Even editing, most obviously in very short time-frames in braunage 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 “spectrom”. Spectral thinking is integral to all sound composition and should be borne in mind as we proceed to explore other aspects of this world.
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 (waveform subdivision: Appendix P6). This will modify all the wavesets to square waves, triangular waves, sine-waves, or even non-sine 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 tapes but complex sounds are just changed in spectral "60Hz" & rapidly changing sine-wave is no less perceptually chaotic than a rapidly changing arbitrary wave-shape. In the sound examples the sound with a word-like attack has waves replaced by square waves, and then by sine waves. Two interpolating sequences (see Chapter 12) between the "word" and each of the transformed sounds is then created (see Appendix P6 & Chapter 12). (Sound example 3.27).

Inverting the half-wave cycles (waveform inversion: Appendix P5) 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 (waveform distortion: Appendix P5) (Sound example 3.28).

We may average the waveform shape over N wavesets (waveform averaging). Although this process appears to be similar to the process of spectral blurring, it is in fact quite irrational, averaging the waveform length and the wave shape (and hence the essential spectral content) in perceptually unpredictable way. More interesting (though apparently less promising) we may replace N in every M wavesets by silence (waveform erosion : Appendix P5). For example, every alternate waveset may be replaced by silence. Superficially, this would appear to be an unimpressive approach but we are in fact thus changing the waveform. Again, this process introduces a slightly ranging "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 waveform in any desired proportions (waveform harmonic dispersion) by making copies of the waveform which are 1/2 in short (1/3 in short etc) and superimposing 2 (3 or) of these on the original waveform in any specified amplitude weighting. With an unnecessary wavesets form the added 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 P5).

We may also remove wavesets in any specified way (waveform blurring : Appendix P5 or reverse wavesets) or groups of N wavesets (waveform renewal : Appendix P5). Again, where N is large we produce a fairly predictable blurring of tone segments, but with smaller values of N the signal is altered in subtle ways. Values of N at the threshold of gross perceptibility are especially interesting. Finally, we may introduce small, random changes to the waveform lengths in the signal (waveform shaking : Appendix P5). 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 descriptively time-ed 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).
SPECTRAL TIME-STRETCHING

A sustained sound with subtle spec and no distinctive onset characteristics may be analyzed to produce a (windowed) frequency-domain representation. We can thus synthesize the sound, using each window to generate a longer duration (phase vocoder). The resulting 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 8x 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 that the human ear perceives the small change from window to window (as fast a "step") as a smooth continuous transition. Once we do too long a synthesise from individual windows, e.g. a 64x time-retch, the spectrum of each window is sustained long enough for us to become aware of the jumps. The continuity of the original source is not reconstructed.

One way around this limitation is to stretch the source x4, symmetrize the result, reanalyze and time-stretch by 4x 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 64x 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 down for us to hear the spectral motion involved. In general we will hear a resulting sound with a gliding harmonic: ("modal") 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 spectral distortions: varieties of a source, zooming in to a maximal time-stretch at a particular point in a source, will produce a particular harmonic artifact in the goal sound. Zooming in to a different point in the source will produce a different harmonic artifact in the goal sound. We can thus produce a collection of richell musical events deriving from the same source (provided enough of the resulting signals are elsewhere similar to each the other). (Sound example 11.12).

The ultimate extension of this process is spectral foreshortening, 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 (grass time-fence) onset characteristics, e.g. an attack-dispersion sound (piano, bell, etc.), even our interpolated spectral time-stretch over a long stretch, will radically

41
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 vibrations to two different groups of partials within the same spectrum will cause the two sets of partials to be perceived independently – the single sound stream will split into two. (See example 3.39).

Spectral fission can be viewed as a number of quite different ways in the frequency domain. Spectral superposition is a process that draws attention to the individual spectral components by isolating, or emphasizing, each in sequence. This can be done by holding a voice over a single pitch by using appropriate vowel formants to emphasize partials above or below the pitch. The composer can apply this process to any sound-source, even if it is in motion. (Sound example 3.26).

Spectral tracing strips away the spectral components in order of increasing loudness (Appendix p.25). When only a few components are left, any sound is reduced to a delay-decelerating decay of (dulling) its constituent. Complex varying resonators produce the most fascinating images as those partials which are at any moment in the permitted gap (the loudest) change from window to window. We hear new partials arising (while others leave) producing "shudder" interest 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 time-wave structure. (See 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 harmonic 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 "virtual" 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. (See example 3.22).

Alternatively we may elaborate the same idea in the time-domain by a process of constructive distortion. By searching for waveforms (zero-crossing points) (Appendix p30) and then repeating the waveforms before proceeding to the next (Waveshift-stretching) we may stretch it time (without altering its pitch see elsewhere for the limitations on this process). (Appendix p55). Waveforms correspond to wavecyles in many pitched sounds, but not always (Appendix p60). Their advantage in the context of constructive distortion is that they only move slowly, having no pitch, have no true wavecycles – but we can still stretch them into wavecycles (Appendix p60).

In a very simple sound source (e.g. a steady waveform, from any oscillator) waveshift-time-stretching produces no artifacts. In a completely evolving signal (especially a noisy one) each waveshift will be different, often radically different, to the previous one, but we will not perceptually register the content of that waveshift in its own right (see the discussion of time-frame in Chapter 1). It merely contributes to the more general percept of motion. The more we repeat each waveshift however, the closer it comes to the grain threshold where we can hear out the implied pitch and the spectral quality implicit in the sound percept by making the indivisible qualitative character of the sound 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 source characteristics by not time-stretching the sound. We can achieve this 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 then increasing as rapidly as we wish to any large value we desire. (Sound example 11.14).

Alternatively, we can use spectral time-stretching to brighten the internal spectral variation of the sound. Time-stretching the onset of the signal, but not the continuation, will alter the sound percept radically (altering the causality) but retain a perceived 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 round these changes by stripping away partials from the time-stretched sound until only the most prominent (say) remain (spectral tracing (see Appendix p28 and Chapter 3). As the spectrum varies in time, partials will grow and leave this favoured set and we will hear out the new entries as "revealed melodies", a type of convolution distortion. If the time-stretch is too short time-varying, these revealed melodies will be regularly pulsed via the time-separation of the original window duration multiplied by the time-stretch factor. Spectral time-warping will create a fluid (accel-decel) tempo of entries and will accelerate into the time continuum when the time-warp 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 stretch the onset of a sound we risk completely altering its perceived character. We overcome this problem by time-variable 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 events. We cannot in this case, therefore, preserve only the beginning of the sound. Ideally we would use an envelope-follower to uncover the linens of the sound and then locate all the onsets, and then apply a time-warping process that left the sound unaltered during every event 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 concatenating them in time, causing the sequence of grains to accelerate, retard or 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 inaudible events, the elements of potential musical phrases. Conversely, time-squeezing 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 "poms" in their own right. (Sound example 11.18).
TIME-SHRINKING

All the above processes may be applied to time-contraction, with qualitatively different results. 

Type-speed variation time-contraction has already been described. Spectral time-shrinking can simultaneously 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 intrinsically lost, i.e., if the grain is now time-dilated once more, the resulting sound will be considerably less time-variable spatially disjuncted than the original. The process of repetitive contraction and expansion will give similar results to the process of spectral blurring (see Chapter 3). (Sound example 11:20).

If grain-shrinks or sequences are spectrally time-shrunk, the individual elements will disintegrate, become less spectrally dense and more click-like as their duration approaches the lower grain-time-frame rate. (Sound example 11:21). In this separation involves the repositioning of grains by reducing intervening silence (where possible), or overlapping or splicing together the existing grains. These latter processes will tend to line grains together. The process hence tends towards tempo-lis like loudness trajectory and essentially toward a continuous perfum, 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-time-frame perfum if grains are deleted from that stream.

In wavelet spectrally time-shrinking wavelet duplication is replaced by wavelet erosion. This process also loses granular or sequential detail in the sound contracts, though in a different way to grain separation contraction. We may choose to view wavelets as a very roughly regular manner (e.g. every fourth or fifth wavelet) in which case the sound becomes increasingly fuzzy, quieter, less detailed. Or, we may choose to omit the least significant (lowest amplitude) wavelets. In this case the sound retains 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.

Finally, we can introduce a some of multiple-source sounds (I) to a sound (e.g. make a single voice appear crowd-like) by adding small random time-changing perturbations to the loudness of the spectral components (spectral blurring). This mimics part of the effect of several voices attempting to deliver the same information. (Sound example 3:10). 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 (broad) into a the static signal (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 locate some music or moving feature of a sound. In a crude way, filters may be used to eliminate unwanted noise or to locate desired sounds, especially digital filters can be very precisely tuned. In the frequency domain, spectral components can be eliminated on a channel-by-channel basis, either as ones or their series locations using spectral splitting to define a frequency band and setting the band loudness to zero, or in terms of their corresponding relative loudness (spectral means) 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 we also "Spectral Fitting". More radically, sets of narrow band pass filters can be used to form a complex spectrum onto any desired pitch 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 sum, 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 partials (e.g. pitch-shifting, add vibrato) and then recombine the two partials perhaps reorganizing the spectrum in a new form. However, if the partials are changed into radically e.g. adding completely different vibrato to each partial, they will not be line when recombined, but we may be interested in the graduated dislocation 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 parts (partial tracking). This is quite a complex task which will involve pitch tracking and partial tracking (as existence where the partial might be) on a window by window basis. Finally it might deal in some way with stereo sound (where the form of the spectrum is not known in advance) and noise sources (where there are, e.g. noise on the partials). This technique is however particularly powerful in as it allows us to set up on additive synthesis model (as we have sound and thereby provides a bridge between unique recorded sound-events and the control available during synthesis.)
Two processes are worth mentioning in this respect. Noise with transient pitch content like water falling in a stream (rather than dripping, flowing or babbling), might be pitch-enhanced by spectral masking (see below). (Sample sound 3.11.) Conversely, all sounds can be assumed 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 3.8 the noise band finally receives 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, slightly bright or dull, grain-like, sustained, evolving, interrupted or segmented) may produce different qualities of noise (see Chapter 8 on Textures). There are also undoubtedly vast areas to be explored at the boundaries of instrumentional and time-fluctuating/spectrumentalism. (Sample sound 3.12.)

A fruitful approach to this territory might be through spectral focusing, described in Chapter 2 (and Appendix II). 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 voice source. The filtered result can vary from articulated noise ensembles (like vocalised speech) following just the fundamental articulation of the original source, to a representation of the partials of the original sound (i.e. half of the original sound itself). We can also modulate fluidly between these two states by varying the analysis window size through time. This technique can be applied to any source, whether it be specifically pitched (harmonics), or infrasonic and gives us a means of passing from articulate noise to articulate noise spectra in a seamless fashion.

Many of the sound phenomena we have discussed in this section are complex combinations of simpler units. It is therefore worthwhile to note that any arbitrary combination of sounds, especially related to music, has a well-defined time-varying spectrum — a well-formed group of tones at a pitch, or a group of tones individually, but simultaneously, projecting their different masses before a concert. At each moment there is a composite spectrum for these events and any portion of it could be geared for the start of sound composition.

SPECTRAL ENHANCEMENT

The already existing structure of a spectrum can be utilized to enhance the original sound. This is particularly important with respect to the onset portion of a sound and we will now discuss more detail of this until Chapter 4. We may reinforce the total spectral amount, adding additional partials by spectral shifting the sound (without changing its duration) (Appendix II) and raising the shifted spectrum on the original. As the digital signal will remain its density precisely, all the components in the shifted spectrum will line up precisely with their non-shifted source and the spectrum will be thickened while retaining its (tonal) integrity. Greater enhancement is the most obvious approach but any interval of transposition (e.g. the amount might be chosen. The process might be repeated and the relative balance of the components adjusted as desired. (Appendix II.) (Sample sound 3.13.)

A further enrichment may be achieved by mixing an already serious spectrum with a pitch-shifted version which is left-right inverted. Theoretically this produces merely a stage–zero (null) sound spectrum but in practice there appear to be frequency dependent effects which lend the resultant sound a new and other spectral "fulness". (Sample sound 3.14.)

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 continuation. We may in this way uncover changing pitch, changing spectral type, changing fragment placement, changing loudness information 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 connectiveness 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 produce less and less surprising perceptual information. (Sample sound 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-sapping) (see above) or simply, perhaps, use the original grain as a superposed onset for time-stretched versions of itself. (Sample sound 11.25.)

Conversely, a sound with continuation will usually have distinctive onset characteristics of grain dimensions. When we time-stretch such a sound we must 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 sound goal, though the continuation of the two sounds may provide sufficient perceptual linkage between them. (Sample sound 11.26.)

As we have seen, grain-strings (and sequence) may be time-extended by granular time-stretching by grain duplication, by granular time-stretching by grain separation or by spectral time-stretching, which later 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 perception of harmonic variation (radiations, acoustic microphones, randomization etc) as the internal pulse of the grain-stream (sequence) changes. In contrast, spectral time-stretching (or Brunnian/harmonics type time-stretching) applied to continuous sounds is more akin to looking at time itself through a magnifying glass — the sound grain longer with no 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 3.) (Sample sound 11.27.)

Granular time-stretching by grain separation of a grain-stream will lead, ultimately, to a "slow" resolution of individual acoustic events which may be re-sequenced in terms of onset-time (rhythm) pitch, or various spatial properties, or any combination of these. Spectral time-stretching or harmonic time-stretching of a grain-stream, if it stretches the grain constraints by a large amount, will reveal spatial qualities and evolving shapes (morphologies) within the grain (or sequence) elements having their own musical implications. (Sample sound 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 phrase, spectral emphases, just as we might produce variants in the English domain, of a melodic phrase. Also, time-warping of a continuous sound can create a kind of timed continuation (see Chapter 5) in the spectral domain I applied to a completely evolving sound evolving sound. Similarly, modulating continuation proportions will become slow glides with quite different musical implications to their underlying sources and we can alter these implications through time-warping. (See example 11.29).

The time-warping of rhythmically paced events provides us with an entirely new area for metrical exploration. It is already possible to work with multiple fixed time-pulses (e.g., mutually synchronized click-ticks in different temps, see Vol 3) or with time-pulse mutually varying in a linear manner (as in the "plucking" pieces of Steve Reich, which arise from new rhythmic, timbral, or spectral parameters through freely controlled delay) but we might also produce internal intervals between rhythmic streams which are themselves acoustically determining, from time to time freezing the streams to pulse-synchroneism in the same manner as was done to the same (or a displaced) pulse-grouping unit (this can be done on two samples).

This device is used in Vol 5 where three copies of an unaltered voice are slightly time-varied in different ways and made to move differently in space. They begin in synchronization, in a single spatial location, and move to a similar place at the section end by appropriate inversions of the speed variations (see Diagram 3). Hence speed divergence and reverberation become elements in the emergence and merging of vocal streams; "comme-torchin" the time-fluid current of counterpart. Note, however, that the sound constellations of two time-varying counter streams do not have to be the same, or even similar. (See example 11.31).

In fact, if the stream constellations are identical, are synchronized at some point and have slightly different time-stretch parameters, their interaction produces fascinating interesting to us in falling from the point of synchronism (see Chapter 2 on Pitch Creation). (See example 11.32). In the previous case (See example 11.33) three precise delay-chip-arrivals were avoided by carefully fading two of the three streams just before each attack would appear.

The interaction of time-varying pulse-streams bears the same relation to fixed-tempo polyrhythm as controlling pitch-gliss dynamics does to musical pitch organization. Time-stretching may lead to spectral fusion (see Chapter 3), e.g. Spectral time-stretching of noisy sounds by large amounts may produce sounds with gliding inharmonic spectra (See example 11.11) – or in conservative dimension as Western time-stretching produces head intensity (See Examples 11.7 & 11.8) and spectral time-stretching reveals window-stepping in spectrally varied sounds (See Example 11.16). In both cases a long time-stretched goal sound may not be perceptually reliable to the source and will be mediating sound structures (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 compression. In particular, phrase structure may be compressed into glides or gliss (or a comprehensible spoken sentence can become a rapid-fire, spectrally irregular sequence of glissis) and compression compressed into the indivisible qualitative percept of grains. Once again, if a perceptual connection between source and goal is required (to build musical structure), mediating sound types may be necessary.

It is possible to extract the time varying) spectral contour from one signal and impose it on another, a process originally developed in the analogues studies and known as "wending" (no connection with the phase wend). For this to work effectively, the sound to be wended must have energy distributed across the whole spectrum so that the spectral contour to be imposed has something to work on. Wending hones vocally on many noisy sounds (e.g. stochastic or sounds which are artificially simplified by adding broad band noise, or subjected to some noise producing distortion process. (See 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 underformalised spectral manipulation in Appendix II.

For a variety of reasons 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 amplitudes to vary as originally. Or we can hold their amplitude stationary, allowing the frequencies to vary as originally. In a complex signal, it holds certain of the amplitudes and, hence the spectral contour, which produces a sense of "freezing" the spectrum when we might have anticipated that the frequencies would create this percept more directly. (See 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, however, a uniform grey area of musical options (or even a few shades of pink and blue) which the music (and for us musicians with noise generation) might suggest. The subtle differences between unknown voices, "g", "d", "p", "k", "b", "f", the variety amongst symbols and simplified glyphs give the lie to this.

Noise can be a matter of degree, particularly as the number of head components in an inharmonic spectrum increases gradually to the point of noise transition. It can, of course, vary from noise-in-wine to noise-in-time, whispered speech in the broad example. It can be moving or less focused towards static or moving patterns, used hand-paste filters or delay (see page 2), and it can have its own complex internal structure. In Sample example 3.9, we hear partamplification inharmonic spectra created by filtering noise. This filtering is gradually refined and the bands become more noise-like.

A good example of the complexity of noise itself in "noisy noise", the type of crediting signal one gets from very poor radio reception tuned to no particular station, from masses of broad-band click-the sounds either in regular layers – cascades – or irregular – masses of breaking twigs or pebbles falling onto tiles – or semi-regular – the gritty vocal sounds produced by water between the toes and palate in e.g. Danish "æ" or from extensively time-contracted speech waves. There are also fluid noises produced by perturbing components, e.g. the sound of water flowing in a wide stream around many small rocks. Some noise shone off into the area of "fuzz" which we will discuss in Chapter 9. (See example 3.9).

These examples illustrate that the rather dull sounding word "noise" hides whole worlds of rich sound material largely unexplored in detail by composers in the past.
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 p.19). (Sound example 3.2).

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

Note that, since stretches produce an ambiguous area in which the original sound appears "colored" in some way rather than getting a new pitch. (Sound example 3.3). Inharmonicity does not therefore necessarily mean multipitchedness. Nor can we have noise with the "kn-ve" example, does it mean bell sounds. Very short inharmonic sounds will sound percussion, like drums, strongly colored drums, or like 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 able 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 formants preserving spectral manipulations - Appendix p.17).

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 long tone. For a vowel to produce, the spectral contour (and therefore the position of the peaks and troughs) must remain where it is even if the period is not stable. (See Appendix p.6).

As we know from singing, and as we can deduce from this diagram, the frequencies of the formants in the spectrum (determining pitch), harmonicity-inharmonicity, notations 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 whistling. (Sound example 3.6).

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

36
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 two dimensional situations 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 transmutation of the percepts. All the perceptible time-varying field properties of the texture-stream will thereby be time-stretched e.g. loudness trajectory, pitch-bend within chorus etc. The revelation of the inner structure of grains may even alter the field percepts (e.g. noise elements becoming inharmonic sounds, or hitches appearing as gliding pitches) of the source sound. (Sound example 11.33).

However time-stretching will produce surprising and unpredictable artefacts when applied to texture streams as the zero crossing analysis will confound the constitution of various distinct grains to the overall signal. With a steady texture, waveform 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).

Breathes techniques with above-granular segments, using regular segment size and zero stretch-duration (see Appendix 11.3) will quickly destroy the separated quality of the texture-stream, as breathage repetitions introduce a "spurious" order into the goal sound. Breathes 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. dynamic change of the hitches 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 stream, then reinserting the original loudness trajectory in exactly the same time-frame as in the source texture-stream. We will discuss this parameter separately 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 overlapping 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 even-out density parameter. To achieve an integrated time-stretch of this sort, any time-varying field properties (hitches field change, loudness trajectory, format change etc.) would need to be similarly time-stretched in the generating indications. We could, however, choose not to alter these features of the stream. In this way, we may create a goal sound which appears less even-out dense than the source sound but not perceptually time-stretched in any meaningful sense. (Sound example 11.37).
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, indefinable property of sounds. Most musicians with a traditional background almost equate "timbre" with instrument type (some musicians producing a variety of "timbres", e.g. piccolo, alto, high bass, etc.). Similarly, in the earlier analog studio, composers "see" into contact with oscillators producing fundamental pitches, noise generators, producing featureless noise bands, and "shred" waveforms which added simple loudness trajectories in three elementary sources. This gave no insight into the audibility 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, due to micro-oscillations or large-scale motion.

HARMONICITY = HARMONICITY

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 grass-size time-frame, the spectrum tunes and we hear a specific (possibly gliding) pitch. If the partials are not in this relationship, and provided the relationship (from window to window) remains 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-oscillations 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 "ka-oo" being gradually spectroscopically elevated (Appendix p95). This means that the partials are moved upwards in such a way that time while number relationships are preserved less and less exactly and eventually lost. (See Diagram 1). Initially, the sound appears to have an indistinct or "sored" around it, a thin, rapidly becoming more and more bell-like.

It is important to understand that this transformation "works" due to a number of factors apart from the harmonic/chromatic variation. As the process proceeds, the tail of the sound is gradually time-attenuated to give it the longer time scale we would expect from an acoustic bell. Most importantly, the morphology (changing shape) of the spectrum is already bell-like. The syllable "ka-oo" begins with a very short broad head spectrum with lots of high-frequency information ("V") corresponding to the initial clash of a bell. This leads immediately into a steady pitch, but the sound formant is varied from "O" to "N", 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 head pitch, persists longer than the higher components. A different initial morphology would have produced a less bell-like result.

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

This suggests the third approach to time-varying. In this case we may generate events at the original density, but for a longer time, thus impose time-dilated field-vibration parameters on the stream. Here the loudness trajectory, the pitch-range variation, the formant changes, the transitions to resonances etc would move more slowly, but the event-fret density would remain as before. This is the textual equivalent of granular time-varying by grain duplication. (Sound example 13.38).

We have begun to touch upon interesting music-philosophical ground. For in this last case, the texture-stream concept is clearly not a "unique" sound-event, in the same way that 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 organized 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-varying a texture-stream. If we include the possibility of spectroscopically time-varying the texture-components prior to generating the texture-stream, we may imagine another option in which the texture constituents are spectroscopically time-attenuated (this time-varying itself changing from constants to consistent, as we proceed through the texture) while (as far as is possible) the temporal evolution of density and field characteristics remains unchanged. This might better be regarded as a time-varying spectro-spectral-configuration transformation of the source texture.

In fact we can time-stretch event-sequences-separation-density variation, sound-duration variation, overall loudness trajectory, pitch-range variation, evolution of the spectral contour (formant-disappearance) etc. etc. 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.
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 (intense spectral crowding, or inner warping or filtering or randomization of perceptual elements etc) we will produce a sound which we recognize 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 upset 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 superposition 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. When distinctions interpolation does make varies is one feeling that we have moved away from one type of sound and arrived at a different type. This concept 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 tenor, 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 (tenor) or it may seek out an ambiguous ground in which two recognition concepts conflict or cooperate within the same experience (the talking beat).

If we now use the analytic data as a set of (time-varying) filters on an input noise source, wherever the analytic window was normal-sized the resultant filters will impose the formal characteristics of the original sound on the noise source (e.g. analyzed 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 formalized- 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 completely 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 systems. This technique is known as comb filtering. Longer delays will give lower and less well defined pitches. (See example 2.2).

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

Phasing or Phongting, 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 8) and this shifted signal is allowed to interact with the unchanged source.

The production of pitch-motion across an appropriate place to end this Chapter as it stresses once again the difference between pitch and pitchness and the power of the new compositional tools to provide control over pitch-in-motion.
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 flute melodic sheet sound, and then imposing a
rapid series of hard-edged loudness impulsions on the melding 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 crystallity have been altered. (Round example 13.1).

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

**MEDITATION : 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 Jungkönige** where the 'pure' pitched singing
voice of a young boy and pure (pitched) voice notes are melded 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-recognizable) sounds, has a metaphysical underpinning (a religious conception of 'unity' in the
context), 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 then
created. Roger Reynolds 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 ong 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, 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
'aura 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 Von 5 the transformation voice—three seems to achieve clear recognition of both sound and goal, and a seamless transition from one to the other without any intervening artifacts which might suggest some other physical-causality, or even hint at the technical process involved. (Sample example 12.2).

Moreover the dynamics of the change is a crucial parameter: in the voice—three example, the voice almost melts slowly into the bee sounds; other transformations in Von 5 are quicker and more forceful, suggesting a generative energy in the sound source, splitting or throwing out the goal sounds—and this dynamics is enhanced by spatialization in the space of spatial notion, or the emergence of stereo images from a mono source. Here, the technical problems are those of unassailability in the spectral transition and achieving the right dynamics, especially as interpretation-mapping often requires a relatively long time-frame in order to work smoothly. Musical context can play a vital role here.

INBETWEENING

The more 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 brains manage to unscramble the many sound impulses on our ears—very time and to sort them into separate sources—We hear mix, not fusion. This is due partly to the ear’s sensitivity to sound synchronicity (or the lack of it) and partly to the persistence of micro-fluctuations of the components in any one source, at the same time being different from that in the other source.

Successful interpolation by mixing can only be achieved if we can define either, or both, of these. In sounds with contrast, the more precise to the sample’s synchronisation of attack (over synchronisation) can achieve an instantaneous fusion of the aural image which is however immediately contradicted by the continuation of the sounds in question. (Sample 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 perceptually 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 wavevector or waveform (see destructive division in Chapter 3). Superpositions of these two sounds in various proportions can create convincing intermediate states (betweenness ? Appendix P46). (Sample 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-no” to “But” from Von 5 (shown example 12.5).

If we began with the goal and source of the sequence and merely mix these in various proportions, we would not achieve a satisfactory set of intermediate sounds. The process of spectral inversion, successively applied, has gradually separated the spectral components (into distinct altered the spectral sound) to a point where they will not simply line together again by mixing.

TETISURRA CHANGE OF UNPITCHED SOUNDS

The techniques of pitch change we have described can be applied to sounds without any definable pitch, like 'artistic sound will, in general, be transposed in a similar way to pitched sounds. However this process will usually change the centre-of-energy (the pitch-basis or 'haut') of the sounds so that they appear to move higher (or lower). (Sample example 2.21). The harmonizer will usually have the same effect. Splitting the spectrum of a broad-based, tone-based sound using spectral shifting may not have any noticeable perceptual effect on the sounds, even when the split is quite radical and churning will often be unrecognizable as the spectrum is already full of 'hauts'. However, the problem of formant shifting when transposing will apply equally well to c.g. unvocal sound sounds. (Sample example 2.22).

We can also use this technique to give a sense of pitch motion in unpitched sounds — noise portamento. (Sample example 2.23).

PITCH CREATION

It is possible to give pitch qualities to initially unpitched sounds. These are two approaches to this which, at a deeper level, are very similar. A filter is an instrument which simulates 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 eliminate parts of the spectrum but, by inversion, to accentuate resonant frequencies. (Resonator filter : Appendix P7). In particular, narrow filters with very high 'Q' (so that the bands allowed to pass are very sharply defined), can slice out narrow ranges of partials from a spectrum. A very narrow band filter can thus impose a marked spectral peak on the omission of sounds, giving it a pitched quality.

In fact this process works best on noisier sounds because here the energy will be divided over the entire spectrum and whenever we place our filter bands, we will be sure to find some signal component to work on. Very tight Q on the filter will produce oscillate-type pitch, whilst less tight Q and broader bands will produce a vaguely ‘sounding’ of the spectral energy. There are many degrees of ‘pitchness’ between pure simple sound vs single oscillator. (Sample 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 noisier towards such chords. (Sample example 2.25).

In a sound with a simpler spectrum, a narrow, light filter may simply miss any significant spectral elements — we may end up with amwl.

A very sophisticated approach to this process is spectral focusing (Appendix 228). 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.
DUAL PITCH

Various compositional processes allow us to generate more than one pitch from a sound. For example, using the harmonizer approach, we can shift the pitch of a sound without altering its duration and then relate this with the original pitch. Apart from the fact that we can apply this technique to any sound, it differs from simply playing two versions on the same instrument because the various and microdistortions of the two pitches remain fairly well in step, a situation impossible to achieve with two separate performers, though it may be closely approximated 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. (Sample example 1.14).

As usual, the harmonizer algorithm introduces significant artifacts 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 one 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 p16). (Sample example 2.15).

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

Small pitch-shifts, superimposed on the original sound will "blur" to the sound, producing the well-known "chorus" effect (as effect produced naturally by a chorus of similar singers, or similar instruments playing the same pitches, where small variations in timing between individual singers, or players, broaden the spectral spread of the resultant modulated sound). (Sample 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 even over a pitch which is moving in a contrary direction. There are not truly two pitches present in those cases but percepts of conflicting motion within the sound can certainly be established. With very careful control, including the phasing in and out of parts at the top and bottom of the spectrum, sounds can be created which get higher in pitch yet lowers in timbre (or lower in High pitch but higher in motion) – the so-called Shepard Tones. (See Or Sonic Art and Appendix p72). (Sample 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 Tibetan harmonics singing). (Sample example 2.19).

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

The difficulties here might be compared with those in the process of in-betweening in animation. Here the direction artist draws key frames for the film, and animates (or computer) 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 skillful 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 use some sort, interleave the data from our two sounds. One way to do this is to switch off analysis windows obtained from the spectral analysis of two sources.

When we produce a spectral analysis of a time–varying sound, we divide the sound into small time–slices, or windows, and analyze the spectrum in each window (using e.g. the fast fourier transform, see Appendix pp.3–4, 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 interchange alternate windows from each sound, preserving the original time–frame, or simply interchange windows as they are, creating a global 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 musical textures. We may also choose to interleave parts (or larger groups) of windows, with a sufficiently large grouping of windows, this will, of course, produce a more richly oscillating between the two source sounds. (Sample example 12.6).

A similar process can, however, be more easily achieved through brausing when we also have control over the segment size and are not therefore tied to regular oscillations. Two or more sources can be teased together (micro–source braising) with control of segment size, segment range, pitch shift, spectralisation of segments etc. just as in single source brausing. This approach perhaps works best in producing a process–focused transformation (see Chapter 1) (e.g. granulated or pitch–spread) in which the two (or more) sources 'per through' the distorting grills of the process simultaneously. The process's artifact grills agglomerate the lack of tone fusion of the sources – both are an integral part of the resulting "mix". (Sample example 12.7).

We can go one step further and attempt to integrate the wave–cycles, or waveforms, of the two sounds. Interleaving wavecycles or waveforms will produce spectrally unpredictable white noise: two interleaved waveforms of different wavelengths will produce a blackboard whose wavelengths in 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 (nonlinear interleaving). (Sample example 12.8).

Similarly, we may impose the waveform (or wavecycle) shape of one sound on the waveform (or wavecycle) length of the other. With simple sources, e.g., square wave to sine wave, this will affect the pitch of the latter on the sound of the former. With complex sources, square waves, the results are once again complex and perceptually unpredictable. Again, we have mutual transformation, but not true transformation (waveform transfer). (Sample example 12.9).
If we have a very rapid sequence or a stream—stream we may achieve a dynamic interpolation between two quite distinct streams by careful element substitution, e.g. we might start with a stream—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 form a broad band noise granule stream onto a single pitch which we might animate with formant—glides and vibra articulation reminiscent of the human voice (granular synthesis: see Appendix). (See Example 12.16).

VOCODING AND SPECTRAL MASKING

The distribution of partials in a spectrum (spectrum form) and the spectral contour (formants) are separable phenomena and impinge differently on our perception. As discussed in Chapter 3 the spectral formants are our sense of the spectral contour corresponding to our format, 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 oboe), we can change the formant variation of the first on the spectral form of the second to create a dual percept (vocalizing). As the spectral contour defining the formants must have something on which to "sit" on the second source, this process works best if the second source has a relatively high spectral contour over the whole frequency spectrum (see Diagram 3).

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 reconstructed by driving a generating signal through these time—varying filters. A sequence of constant squarewave type basis of appropriate pitches (for voiced vowels and consonants) and static white noise (for noise consonants and unvoiced speech) played through the time—varying filters can be used to reconstruct the original speech. The process can also be used to reconstruct the speech at a different pitch (change the pitch of the basis), or speed (change the rate of succession of the filters and the harmonics), 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 vocal). It will usually not have a form, or even a stable spectrum, but we can enhance the formant transfer process by "whitening" the second source, i.e. by adding noise directly 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 vowel and a second source and is often called vocoding. It should not be confused with the phase vocoder. (Similar sounding procedures may, however, be attempted, using spectral data extracted by phase vocoder analysis).

This type of interpolation might also be applied progressively so that 'voices', or lack of it, emerges out of a continuing non—vocal sound. (See Example 12.11).

Making the aha '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 in which an appropriate wave—breaking—shape loudest—pulleys is applied (using a chord conductor!). If the spectral type of the sounds within the entire stream is made to vary appropriately (e.g. lack of clenches initially, 'I' and 'A' clenches at the wave peak, 'Y' clenches 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.

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 pitch—transposition. As this does not change the window duration, the pitch is shifted without changing the sound duration. This is spectral shifting. (See Example 2.13).

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 p16).

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 modified spectrum on the newly shifted partials. The four stages might be as follows...

(a) Extract the spectral contour using linear predictive coding (LPC) (Appendix p12—13).
(b) Extract the partials with the phase vocoder (Appendix p11) or with a fine—grained LPC (see spectral—convolving below).
(c) Flatten the spectrum using the inverse of the spectral contour.
(d) Change the spectrum.

(e) Reinstate the original spectral contour (See Appendix p17).

Ideally this approach of separating the formant data and the partial data should be applied even when mainly imposing voices on a sound (see Chapter 10) but it is computationally intensive and, except in the case of the human voice, probably unnecessarily faithful 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 characterized often by a single soundhole (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.
**Diagram 3**
Replace each wave set by three shortened copies of itself. Wavelength reduced to 1/3, therefore frequency x 3, hence transpose up by interval of a 12th.

**Diagram 4**
Take wave sets in groups of three. Replace each 3-set by 3 shortened copies of itself.

**Diagram 5**
Find true wave cycles, with help of a pitch-tracking instrument. Replace each wave cycle with 3 shortened copies.
Another way in which we can achieve interaction between the spectra of two or more sounds is via spectral mixing. Here we construct a goal sound from the spectra of two (or more) source sounds by selecting the loudest partial on a frequency-bands by frequency-bands 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 modified sources will be suddenly revealed if the first source passes, or gain quieter. Hence aspects of the spectra of two (or more) sources may be played off against each other in a componential interaction of the spectral data. This technique should, in general, be regarded as a form of spectrally interactive mixing, rather than interpolation in the time sense. However, if one or two (or more) source sounds have really pitched glides which are strongly pitch-related, alterations in the loudness balance of source sounds will produce interpolated spectral glides. With glides, or other sources with variable formants, this interaction will be particularly potent. (Sound example 12.12.)

SPECTRAL INTERPOLATION

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

In Appendix G2 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 synchronization between the two sounds, where the time-step corresponds to the window duration.

We can now apply a process of moving the amplitude (audible?) 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 these values in sound 2 in window N+1, then in windows N+2 etc., the middle value in window N+1 will be 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 (base) value in window N of sound 1 towards the ultimate value (sound 2 in window N) of sound 2. The latter, 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 'weakening' of one spectrum into the 'weakening' 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-structures within each sound are mentally synchronized 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-structures themselves. (Listen to Sound example 12.2.)

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 adapting this approach for every sound we use, would make sound composition unfeasible. 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 studio work this is 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 changes 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. portamento) this 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 to specify a change in terms of its final velocity (tape acceleration). The availability of time-variable processing lets us a new order of compositional control of such time-varying processes. (Sound example 2.18.)

Wavetable resynthesis 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 waveform (in the sense of a part of zero-crossing : Appendix G9) is replaced by a shortened copy occupying the same time as the original one. (Diagram 3 and Appendix G11.) This technique is very fast to compute but often introduces strange, signal-dependent (i.e. varying with the signal) artifacts. (It can therefore be used as a process of constructive distortion in its own right!) Grouping the waveforms 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 is chosen again depends on the signal.

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

The technique can also be used to transpose the sound downwards, replacing N waveforms or waveforms 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 G11 : Sound example 2.11a.)

A more satisfactory time-domain approach is through resynthesis. To lower the pitch of the sound, we cut the sound into short segments, then slow them down in tape-speed variation, which lengthens them, then gel 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 interp segment which would lead to unwanted echo effects within the pitch-changed sound. This technique, used in the Harmonizer, works quite well over a range of one octave, up or down, but beyond this, begins to introduce significant artifacts: the signal is transformed as well as pitch-shifted. (Sound example 2.12.)
There are also some fact techniques that can be used in special cases. Once the pitch of a sound is known, we can estimate its pronunciation using voice–fingertip representation (Appendix P36). 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–intersection) the odd harmonics while reinforcing the even harmonics. Thus, if we start with a spectrum whose partials are at 100, 200, 300, 400, 500 Hz, with a fundamental at 100 Hz, we are left with partials at 200, 400 Hz whose fundamental lies at 200 Hz, an octave above the original sound. A modification of our method, using the filterbank exemplars, allows us to make an octave downward transposition in a similar fashion. The process is particularly useful because it does not disturb the contour of the spectrum (the formants are not changed; 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 pitchshifts. 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 pitch values at any point. For example, the subtleties of portamenti and vibrato articulation in a particular established vocal or instrumental idiom, or in a naturally mutated 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 any stage having a quantifiable (immeasurable) 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 on hand, or devices that modify collections of similar objects (strings, bowed bars, wooden bars, etc), or with variable access to the same objects (true fingerboards, 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 at steady pitch (or pitch set)). Most important is the way in which the pitch of sound cannot be reproduced (breaking a sheet of glass ... every sheet will break differently, no matter how much care we take), or because the density of these sound cannot be precisely remembered (a spoken phrase can be repeated at different pitches by the same voice, without a narrow range but, of course, natural sound inflections, the fine details of articulation cannot usually be precisely remembered).

In fact, changing the pitch on an instrument does involve some specific complications, e.g., low and high pitch on the piano have a very different special quality but we have come to regard these differences as acceptable through the skill of the performer (and the change of the piano strings causes in the same sound–box and there is a relatively smooth transition in quality from low to high strings), and the fallibility 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.

There are a number of refinements to the interpolation process. We may interpolate amplitude data and frequency data separately (at different times and at equal different times) and we may interpolate in a linear or non–linear fashion, and we must choose these parameters to a way which is appropriate to the particular pair of sounds with which we are working (see Appendix P30).

There are also perceptual problems in creating truly seamless interpolation between two recognizable sources. First of all, source recognition takes time and we need enough time for both source and goal to be recognized, 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 transitional transition is to be achieved without intervening artefacts which either suggest a third and subtly blended individuality in the sound, or even reveal the mechanics of the process of composition itself.

A more difficult problem is created by cursive writing in perception, e.g., in the transformation voice–object, we tend to perceive: "a voice = a voice? = a voice? = it is here! = in a sudden switch = we recognize the goal percept. It may be necessary, to achieve a truly seamless transition, to create inaudible "fake trails" which distract the ear's attention sufficiently in the very point of maximum uncertainty for the transition to be accepted. In the voice–swell transition in Yor 3, a very high–register, low–level part of the spectrum (very low a note of maximal depth) is heard (a way which is not consciously registered but seems sufficient to assume the sudden cursive switching which had not been overcome by other means.

This type of special 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–swelling (so that certain levels sound and stage sound the find through the interpolation) and we place no aesthetic restrictions (such as the goal of "smoothening" in the examples above), on the interpolation sound–types, it should be possible to produce a set of spectral transitions between any two sounds with one word of caution! Equal small changes in spectral parameters, sound not lead to equal small perceptual changes. In fact, a slight deviation in special form (e.g. harmonics) 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 slice of interpolation along a set of sounds may require a highly non–linear sequence of processing parameters 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 interrelations. In the classic stereo, imposed reverberation can help to give a sound–perception for space impression of "spatial integrity" (this sound was apparently produced in a single place in a given space). More profoundly, spatial streaming will lead to separate a fixed image, in a sequence or strict–stream, if set of element moves left and the right, we experience aural stream disconnection.

A movement from move to stereo can, however, be used to enhance, or 'deemphasize' the process of dynamic spectral interpolation. In Yor 2 of the voice–other interpolations start with a voice in moves at front center stage and interpolate to a stereo image (crowd, see etc) which itself often moves off over the listeners heads. Spatialisation and spatial motion hence reinforce the dynamism of the transition.
STORING VALUES OF A TIME-CHANGING PITCH:

- Store at regular times: Accurate but inefficient.
  (Stable pitch in segments 'a' stored many times).
- Store at regular pitch-steps: Efficient but inaccurate.
  (Segments 'a' replaced by slight perturbations).
- Store at regular gradient change: (Gradient is ratio of change of pitch).

Efficient and accurate.

PITCH TRANSFER

In each analysis window, multiply all values by transposition ratio for this window. e.g. when transposition ratio is 1/2...

Derive (time-varying) transposition ratio from the pitch-trace of the original sound and the trace of the desired pitch.

pitch as frequency

Original pitch

Transposition ratio

Desired pitch

pitch as frequency

Time

Time
We can also attempt pitch-tracking by partial analysis (Appendix 7). This means the frequency-domain we would expect to have a sequence of (issues) windows, in which the most significant frequency information has been extracted in a number of channels (see in the phase vectors). 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 may then separate the true partials from the other information by looking for peaks in the results.

Then, in 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, each within a grid of prints of a frequency, the number of partials frequencies will have a highest common factor, e.g. partials at 100.5, 201, 307 and 313.5 have an HCF of 0.1. We must therefore reject almost all values. A good lower limit would be 16 cycles, the approximate lowest 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 2). In such a relatively straightforward case, pitch-tracking can be very accurate, with perhaps occasional oscillation problems (the pitch can be assigned to the wrong octave). However, in the general case (e.g., speech, or synthesized sequences involving harmonic 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 hashpolar table (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 the table? More precisely, if a pitch is changing, when is the rate of change adjusted to have changed? (See Diagram 1).

If we are working on an Lp2pitch reference frame the task is, of course, much simpler. If we do not confuse ourselves with frames, to be completely rigorous we could have the pitch value found at every window in the frequency always representation, but this is reportedly wasteful. Best to decide on our own ability to discriminate rates of pitch motion and to the pitch-detection instrument a point-to-point change threshold which, when exceeded, means a new value to be recorded in our pitch data file.

**PITCH TRANSFER**

Once we have established a satisfactory pitch-track 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 (obtaining) pitch-trace, either directly, or from a second sound. By comparing the two traces, a pitch-following instrument will determine the ratio between the new pitch and the original pitch at a particular window time (the instantaneous transpositional ratio), then multiply the frequency values in each channel of the original window by this ratio, hence altering the perceived pitch in the synthesized sound. (See Diagram 2).

**CHAPTER 13**

**NEW APPROACHES TO FORM**

**BEYOND SOUNDS-OBJECTS**

There is still the danger of regarding sound-composition as a means to provide self-contained objects which are disadvantageously to be controlled by an external architecture of pitch and rhythm along traditional lines. In particular, MIDI technology (1994) makes this version so easy and other approaches so non-digital that it is easy to give up this stage and revert to purely conventional 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, an intuitive control, large-scale musical forms in general, 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 last Twentieth Century, will inevitably be limited. I would particularly stress a detailed appreciation of sourcing styles and rhythms of declamation from around the world's cultures as a way to appreciate melodic-harmonic structure and the rhythmically written. Similarly, a study of the World's languages reveals the great range of sound materials that enter into everyday human sonority organisation somewhere on the planet.

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

**MULTIDIMENSIONALITY**

Given the priorities of modern Western Art Music, and the fact that these priorities are constructed into the instrument technology (from the tempered keyboard, or keylist, to, the MIDI protocol) it is easy to view music as a two-dimensional (horizontal/vertical) 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 relations sets in terms of the pitch and pitch-field, or their pitch and pitch-field, their vowel set, harmonic content, note density, vowel-accentuation-type etc. etc. with each of these treated as distinct perceptible organisable parameters. The of these parameters are perceptually separable in every case, e.g. vowel accentuation type may be independent when sound density is very high. But all definite sonic situations we have sound parameters to dispose.

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 interoperation between quite different means (e.g. "Wax-" or "Bell"-In sound example 12.5). This latter sound interoperation illustrates the fact that we can
certain scales of perception between definable limits without paraproducting 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 stpped line in any direction.
We may thus explore this multi-dimensional space establishing relationships between sounds (diachronically or through different properties unto which are both perceptible and musically potent in some sense.

Given this multi-dimensional space, we can generalize the notion of "modulation" (in the literal sense). We are already aware that recomposing Htpich music can fundamentally change its affective character. Music transformed by the instrument, or within the instrumental context, a change in "expression" in the production of the sound stream (e.g. bowing attack, audible vibrato continuation on violins.

The distinction made between "transmit" and "exposition" is in fact an arbitrary, ideological divide. All the changes to the sounds can be traced to conceptual properties of the sounds and the control of those structural properties. The distinction structural/expression arises from the arbitrary divide created by the limitations of notation. Features of sound we love, in the past, been able to sustain with some exactitude (Htpich, duration) are opened up to "structural" control (or at least rationalized expression) by composers and performers. Those which remain invisible or vague in the notation are out of reach of this rationalizable 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 distorts the basis for this descriptive view. The multi-dimensional complexity of the sound world is opened up to compositional control. In this new space of possibilities, means (or rationalization), must come to terms with limitation. 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 such parameters as a separate entity.

We may group properties into related sets, or link the way we perceive properties with the variation of another (e.g. vibrato speed with duration) 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 when 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 (tone, accent, information, questioning etc.) which can be conveyed by multi-auditory articulation of the sound space, is something we take for granted in everyday social interactions and in theatre context.

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

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. (See example 2.4).

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

Returning now to portamento and considering rising portamento, it is possible to relate such moving pitches to Htpich if the portaments have special properties. This, then, if they are coter-weighted and the ones focused in a fixed HAmusic field, we will assign Htpich to the events, regarding the portaments as ornament or articulations to those Htpich. (See example 2.5a). Similarly, end-focused portaments (often heard in popular music singing styles) where the portament settle clearly onto values in a HAmusic field, will be perceived an ornamental ornament to Htpich. (See example 2.5b).

Portaments too no such weighting, however, will not be perceived to be Htpich. (See example 2.5c).

The same arguments apply in falling portaments and even more powerfully, to completely moving portaments. (See example 2.6). Nevertheless all these sounds are pitched in the spectral sense.

Using our new computer instruments it becomes possible to follow and extend the pitch from a sound event, but this does not necessarily (or usually) mean that we are assigning an Htpich, or of Htpich 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 Htpich or scale, mode or HAmusic field existing within our thinking, or our perception. This should be borne in mind while reading this Chapter as it is all too easy to mislead in leap-leap-into thinking of pitch as Htpich.

A good example in traditional musical practice of pitch not created as Htpich can be found in Xenakis' Pithagoras where portaments are organized in ornamental fields governed by Polus' formula or in portaments of portamento, which transform rise and fall without having any definite Htpich. (See On Sonic Art). (See example 2.7).

In contrast, Paul de Nolles uses pitch extraction on natural speech recordings, subsequently reinforcing the speech pitches on synthetic instruments so the speech appears to ring. (See 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 recognize 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 2.9).

104

27
what it bears. The odd harmonics, say 500, 500, 700, 900, will continue to vibrate a fundamental of 100 cycles but will take on a chord-like quality (chiaroscuro produces only the odd harmonics in the spectrum) and move into one of the loudspeaker. The remaining harmonics, 200, 400, 600, 800, will be interpreted as having a fundamental at 200 and 800 and hence a "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 3:2).

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. As "A" and "C" played on two flutes (or on one flute) are two distinct sound events, each having its own integrated spectrum. Each spectrum is integrated with itself because its internal microstructures are summed in parallel over all its own partials. But these microstructures are different to those in the other spectrum. Within a single spectrum, however, partials might roughly correspond to an A and C (parallel but in exact numerical proportions, unlike in the tempered scale) but will also have exactly parallel microstructures and hence fuse in our sensory 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 patently quite distinct.

To avoid confusion, we will try to reserve the words "Harmony" and "Harmonic" (capitalized as shown) to apply to the traditional concepts with relations among notes in Western art music and we will refer to either a spectrum having pitch as a pitch-spectrum or as having harmony, 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 constraining term to "intensiveness".

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 spectrum is pitched in the spectral sense. It is difficult to speak of a spectrum having pitch, as we do in a different sense of conventional harmony. This sense of "having a pitch" is being able to assign a pitch to a specific class like E-flat or C#2 is quite a different concept from the inteml concept of pitch described here. We will therefore refer to that traditional concept as Haptic, an abbreviation for pitch-as-sensation/intensiveness/harmonic.

Perception of pitch depends on the existence of a frame of reference (see Chapter 1). Even with steady (non-oscillating) pitches, we may still have no sense of Haptic if the pitches are selected at random from the continuum of voices, though often our cultural predispositions cause us to "hear" the notes onto our pervecntion frame of which 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 approximative to a Harmonic field and finally the 'haptic' set locked onto that Harmonic field. (Sound example 2:3).

THE STRUCTURAL FUNCTION OF INTERPOLATION

It is instructive to examine the form of some sound composition pieces to illustrate this multi-dimensional approach. The sound examples here are from Janssens of Fire (1952-54). In Sound Examples 13:1 we begin with a vocal sound whose tail is gradually t音-struck to a gentle remnant at its end. The prior context is that of voice sounds, Voice sounds themselves are only recognizable as such through a complex interaction of properties (pitch-vibrancy, pitch-glissade speed and range, formant and formant-glissade set, note 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 to vibrato or tremolo articulated in numerous ways. Each event in this segment begins vocally, but the spectral-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 tremolo-like ideas, but now the loudness trajectory of the tremolo sets so deep that the originally continuous sound breaks into a succession of voice-like units. 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 pitch-darkening 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-extended version of the tail occurs without the vocal instance, thus understating the sonic "key change" which has taken place. As this event closes, the wood-like events rise in pitch and density, establishing a high granular texture. The voice has now 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 events.

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

In my description of this event I have tried to emphasize the structural importance of similarity by comparing to the 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 another discrete cyclic system - but, by contrast, it is wonderfully multi-dimensional, and we still retain some notion 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 find 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 inverses), the decelerating (and related accelerating) rhythmic patterns. What is significant is that many different dimensions of sound organization can provide structural reference points. We need not have the traditional hierarchy: pitch, 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 pitch event is inserted. As the sequence proceeds the units are given spectral pitch through successive filtering and material focus partly due to the pitch domain. Once the "bow" enters, the rhythmic pattern is lost in a dense texture, but the pitch-focus remains in the harmonic field of this texture. In the foreground the "bow"arks" are linked by a new device, the falling portamento-shape whose origin we hear when the spectrally-time-structured voice peers through the texture.

In Sound example 13.3 we begin with a vocal texture increasing in density, which begins to rise in intensity. As it does so the texture wherein the line produces a vocal wash which is subsequently filtered to produce a multi-pitched inharmonic sound. The rising intensity of the texture is felt perceptually as a secondary property (an articulation). But it soon becomes the binding structural form for the ensuing sequence, which glide up and down.

It is the very multi-dimensionality of the sonic continuum (and in some sense forced to call upon different perceptual foci as binding elements) different times. I am not suggesting that one consciously concedes in the music that an intangible sense of formal cohesion over an inarchitectural 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 scales and barlines of the musical scene reinforces a culturally received conception of the musical scene as lying on a grid, or lattice, of discrete values, partially 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 units: finite sets of pitches and the associated intervals, over durations as multiples or divisions of a regular set (e.g. a chord). The grid holds from the reality of the musical continuum from which it is carved, but yet as the integers and the rational numbers (functions) are only special cases lying along the underlying continuum of the real numbers, so pitch sets and countable rhythmic units are special cases of an underlying continuum of frequency and duration values marked by the continium of format types, notational forms etc.

Rationality, however, is in the ear of the beholder. For music is an acoustically grounded 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-tons are equal and the fifth is simply 7 steps up a 12 sem scale. Musicians from other cultures, WHAT IS PITCH?

Certain sounds appear to possess a clear quality we call pitch. Whereas a symbol or a sound—tone 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 periodicity 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 may not in fact be expressed in the notation of "harmony" in traditional sound. What happens when the periodicity are not in this simple relationship is described in the next section. Thus the numbers 200, 300, 400, 500 all imply 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 sound example 13.2 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 note 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 the 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 "contour." This feature is best illustrated by looking at examples.

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 loudspeakers. This happens because a phenomenon known as aural steering. When sounds from two different sound sources enter one ear simultaneously we need some mechanism to disengage the partial 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 psycho-acoustics (pitch) in air, or in water, and loudness which all naturally occurring sounds exhibit. The pitches derived from one sound will all join in parallel with one another, while those from the other sound will just diffuse directly has 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 experience, however, we have removed this clue by maintaining synchronicity of microfracture between the pitches coming from the two loudspeakers. Hence the ear does not unscramble the data into two separate sounds. The voice remains a single percept. (Sound example 2:1).

If, now, we gradually add a different vibrato in each of the partials, the sound image will split. The ear is now able to group the 2 sets of data into two aural views and assigns different source sounds to
In sound composition, a relationship between two sounds is established only through nearly perceptible similarity or relatedness, regardless of the methodological nature of the process which transforms one sound into another.

Another important aspect of Saravat’s work was its distinction between sound-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, including the envelope) and process-focused transformation (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 pitch and motion). 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 permutation 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.

Transformation-focused in the score, however, means the same infinite potential that the infinity of natural sound sources offers 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.

and dozens of various pre-empted approaches to taming (e.g. just intonation), however would beg to differ from this analysis.

For them the interval we knew 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 3 raised to the power of 1/12, or approximately 1.998397077. 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’ 5th).

The integers in the tempered scale, roughly corresponding to the medivial ‘des’ in music’, is even more instructive in this regard. For the tempered scale tone corresponds to a precise frequency ratio of the square root of 2. 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 spurred 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 mathematics, 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 hopeless 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 infinitesimally small part) of the Real.

In a sense, Western musical practice has remained overseeable 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 composer 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 increased 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, creates for us a continuous sonic domain from what are fixed and countable values. The composer thus provides a link between the world of rational
mathematics and the continuum of sonic experience (re-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 "finite" ratios in the sense of tuning systems, or temporal properties, is seen for the idealism that it is. We come face-to-face with the final, 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 temporarily over the continuum) in performance practice, is now recordable, recognizable, analyzable, measurable 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 "intervals" 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. Among all Western Art Music as it appears in the notation is concerned with articulating sequential forms.

The invention of the orchestral crescendo by the Manhattan School of Symphony 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, 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-translation. (More recent notational developments include special 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, decel, poco a poco, etc.) or rapidly ( molto al acecel). Interval articulation of the rate of change is not describable through some twentieth century composer, like Schoenberg, have attempted to extend notation practice to do this, in works for solo performer). Performance practice may involve the subtle or involved use of "rubato" (performance "left" 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 strict time information. But the fact is, we cannot notice this kind of subtle tempo fluctuation. Does this mean therefore, must be abandoned as an element in musical construction?

The subtle articulation of pitch, continuity and spectral format properties can be observed in the vocal music practice of many contemporary musical traditions, as well as written jazz and in popular music. Music = notation, at least partly, by our tradition, can carry the morphic form information from one generation of performers to another. It does not get tamed off into a separate domain of "performance practice" separated from "stable-property musicology and subsequently devolved.

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 sound may also vary in time (in subtle and complex ways. In general, any parameter we give to a process (pitch, speed, loudness, filter contour, density etc.) should be representative 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 depict in a notation system are turned into subtly varying foyer by musical performance. Where this is not the case (e.g., quantized sequencer music on elementary synthesis modules) we quickly lose the flexibility of the musical stringency.

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 lack of musical validity. In a similar fashion an "effects unit" used as an all-in-everything process on a voice or instrument, is simply that, an "effect" which may occasionally be appropriate (to create a particular general acoustic ambiance, for example), but will not sustain our interest in a compositional manner 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) vibration, applied to a fairly static sound, gradually giving it a vocal quality; a gradual spectral reshaping of a long sound causing its pitch to split into an inharmonic tesserae, and so on. Transforming a sound does not necessarily mean changing it in an all-or-nothing way. Dynamic spectral modulation 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 multistability of sound manifests another significant musical issue, discussed by Alain Savouré at the International Computer Music Conference at IRCAM in 1985. As we can do anything to a signal, we may decide to consider whether the source sound and the goal sound of a compositional process we in fact are perceptually related, or at least whether we can define a conceivable 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 exploited and who's results can also be seen as a source, by definition defines a musical relationship. This traditional conundrum is out of the question in the new musical domain. An instrument which replaces every half-waveform by a single polar equal in amplitude to the amplitude of the half-waveform, is a perfectly well-defined transformation, but reduces most sounds to a uniform cracking noise. Nevertheless, for complex real-world sounds, each cracking signal is unique for 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.

108

23
DIAGRAM 2

Stable pitches defining a reference frame.

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

DIAGRAM 1

Conventional crescendo.

Stepped crescendo.
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 'true' performance category in Western Art Music and the internal fixation of analyses and evaluation, it is difficult for music analytic thought to take on board morphic forms. Such a vast literature of music analysis exists, grounded in sequential properties. By contrast, there has been no means to resist, let alone describe, and analyze, morphic form. The opening 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 faced by musical texts, however, this area will remain a closed book.

From a perceptual perspective, even in the most elementary case, a motion from static A to some B, we are able to distinguish articulations of the motion. For example, a pitch–parameter which rises steeply before slowing down on the goal pitch, can be differentiated from a linear parameter 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 droning/oscillating portamento which accelerates away from a pitch, a "leading-from" or dispersive morphic.

2. A rising portamento which slows as it approaches the final pitch 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 accentuate.

3. A sense of morphic stability achieved on a sustained tone with static or 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 cycle), formant motion, harmonicity shift, density fluctuation etc., etc.). In other cases, simultaneous morphic develop of parameters may lead to contradictory directions. A tone whose vibrato–rate goes gradually but whose vibrato depth simultaneously widens comes 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) pitch–field perception to pitch–glide field perception.

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

We can see the same mixed classification in the Indian raga system where a raga is often defined in terms of both a mode (of static pitch values) and various metric figures often involving sliding interactions (suction-types). In general, in the case of pitch reference–frames, values lying off the reference frame may only be exact as ornamental or articulatory features of the musical stream. They have a different status to that of pitches on the reference frame (cf. "blue note" in jazz, certain kinds of harmonic 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 (though the frequency of a pitched sound produces the same pitch, one octave higher, pitch reference frames are cyclic, repeating themselves in each octave over the audible range. In contrast, in the vowel space, we have only one "act" of possibilities but we are still able to recognize, without reference to other vowel sounds, where we are in the vowel space. We have a sense of absolute position.

People with pitch which can deal with pitch space in this absolute term, but for most of us, we have only some general conception of "up and down," we can, however, determine our position relative to a given pitch, using the notion of interval. If the octave is raised by a mode or scale into an symmetrical set of intervals, we can tell where we are from a small set of notes without knowing the key note because the interval relationships between the notes orient as within the set–of–intervals mapping up the scale. We cannot do this, however, with a completely symmetrical scale (whole–tone scale, chromatic scale) without some additional clue (See Diagram 8).

Cyclic repetition over the domain of reference and the notion of interval are specific to pitch and tone reference–frames. However, tone reference frames which enter our perception of rhythm are particularly resonant in musical time and I will propose further discussion of these until Chapters 9.

Traditional Western music practice is strongly wedded to pitch reference frames. In fact on many instruments (keyboards, fretted string) they are difficult to escape. However, in some 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 change the reference frame itself, through time, or move on to, and away from, such a frame. Computer control permits the very precise exploitation 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 -- that we already find in tone languages in China and Africa.
in the sequential domain. Stream interaction in an elementary sense can be heard in Steve Reich’s phasing pieces and in the solution section of Vice–5 (see Chapter 13) but in neither of these cases is the material in the separate streams undergoing real morphic change (i.e. 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 tonally and spatially at a later time.

Imagine also a tight situation in which we see the quite separate evolution of two characters with entirely different, but internally consistent, goals and desires. Then external circumstances compel the two to meet briefly in a small room. The theatrical outcome is dramatic and we, as audience, have an instinct of how this collision of personalities might turn out, whereas the characters are unprepared for the change encounter. Also 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 versatile composer who can grasp the concepts and mathematical controls of morphic flow process.

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 calculated remains of the Nazi’s (e.g. the shell of the Nazi) tell us about the processes of continuous growth which formed them, into any spatial proportions we may observe in the final forms are merely the traces of these continuous processes, time-flows revealed in space.

What unites 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 realized through a process of branching flow. The overall state of a particular flow pattern (as result being the current state of the tree) links our perception of one tree to another in the same species.

Unlike the growing tree, a musical event moves from its time-flow evolution in a spatial object, it is intrinsically evocative and a musical score is not the musical experience itself, but a set of instructions to generate this evocative substance, just as the genetic code is a set of instructions to grow an organism. The symmetries of the DNA nucleotides are one, 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 a codify experience, professional practice, social order and even human relationships in strict categories. Others see categoric divisions as temporary in which we attempt to control 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 ideal ideals to which everyday reality corresponds with varying degrees of success. In such a world, the concept of *limen* or the compositional limit or method, may be viewed as a musical ideal and its manipulation in performance an adequate or poor reflection of this ideal object. Reality is set in stone, or print, as in the exact proportions of architecture, the unchanging flux 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 mental attempts to hold this flow in our grasp so we can understand it and perhaps control it in 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 these tests 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, is seeking the measure of the evanescent flow of sonic events, attempts to grapple with the very essence of human experience.

---

**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 flare, etc). Furthermore, motion may itself be in motion, e.g. cyclically varying pitch (rubato) may be accelerating in cyclic-duration, while shaking in pitch-depth. Cyclic motions of various kinds (sine, vibra, spectral vibra, 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. Yet separation is complicated in Western Art Music by the use of a diacauseuse notation system which assumes 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 indirectly reproducible. We can, for example, reproduce the direction, duration and range of a rising pentagram and in practice, we can differentiate a single-weighted from an end-weighted pentagram. (See Diagram 6).

In many non-Western cultures, culture 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 even from immediate memory. The difficulty of immediate reproducibility makes reproduction in performance very difficult and therefore acquaintance and knowledge through humbly impossible is acquire.

With computer technology, however, complex, time-varying aspects of a sound-event can be tied down with precision, reproduced and translated to other sounds. For the first time, we have sophisticated control over sonic motion in many dimensions (pitch, loudness, spectral coherence or formant evolution, spectral harmonics/phaseharmonics 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!) organize a sound property in relation to a set of reference values. These provide reference points, or markerless (inc. in the continuum of possible possibilities. Thus, for example, any peritonal language provides a phonetic reference-frame distinguishing these vowels and consonants (up to be regarded as different (and hence capable of articulating different meanings) within the continuum of possibilities. These distinctions may be subtle ("U" 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. Conventions like "W" and "Y" (in English) and various vowel diphthongs are in fact defined by the motion of their spectral contours (formants: see Appendix 4). In general, reference frames for motion types are not so well understood and conventionally do not serve strongly in traditional Western art music practice. Nevertheless, we are equipped in the sphere of language perception, at the level of the phonemes ("w", "y", etc) and tone languages like Chinese and "tone-of-tone", to make very subtle distinctions of pitch motion and spectral motion. They are key to our comprehension at the phonetic 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.
Thus, for example, we may spectrally time—space and change the hardness trajectory (envelope) of a vocal sound, producing wood-like attacks, which are then progressively deterred to sound like simple pitched 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 grow—a xylophone note, a drum-stroke, a vocal click. It may be inharmonic, but permitted to resonate through an associated physical system, e.g., a cello soundbox for a plucked note, a resonant wall for a drum stroke. Or the material itself may have internal resonating properties (hills, grooves, metal tubes) producing a periodically attenuated continuum of sound.

In the case of continuous excitation of a medium, the medium may resonate, producing a steady pitch which varies inharmonically with the energy of the excitation e.g., in a violin. The medium may vibrate at a frequency related to the excitation force (e.g., a note—driven string, 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 tinted sound (called "E", drum roll etc).

The vibrating medium itself may be classically mobile—a string, a fixed metal sheet, a mammalian larynx—so that the pitch or spectrum of the sound varies through time. The material may be only greatly caused into motion (the air in a "scream", the shell in a Rainstick) giving the sound a soft onset, or the material may be loosely bound and granular (the sand in a Shaker or Wind—machine) giving the sound a diffuse continuation. Resonating systems will stabilize after a variety of transients or unstable initiating events (flute breaths, coarse hammer blows to a metal sheet) so that a complex and disconnected onset leads to a stable or slowly evolving spectrum.

I am not suggesting that we consciously analyze 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 hear in mind in sound composition.

In a similar and not easily dissociated way, the onset (or attack) properties of a sound give us some idea of the cause of that sound—a physical blow, a striking event, 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 "cessation" in the onset can adhere to the ensuing sound in the most unlikely cases.
with coarse pinnas) where the sound is clearly made up of many individual randomized attacks. However, 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 lengths, we reach the sphere of the Power. 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 cuts or boundaries, but questions of the interpretation of our experience. A still, without variation, being over four bars may be regarded as a note-articulation (an example of continuation) and may exceed in length a metric phrase. But if still with a marked series of business and power 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 (consonants and stills forming 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-slid it to become a sound of segmental morphology (see Chapter 3). 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 slide 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 sound-time-frames up to the level of the duration of an entire work. These extended time-frames are the basis of our perception of both rhythm and larger-scale form and this is more fully discussed in Chapter 6.

THE SOUND AS A WHOLE – PHYSICALITY AND CAUSALITY

Most sounds larger than a gram can be considered in terms of an onset and a continuation. A detailed description of the typology of sounds can be found in On Sonic Art and in the writings of the Groupe de Recherches Musicales. Next, I would like to draw attention to two aspects of our aural experience.

The way in which a sound is attached and continues provide 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 sound 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 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 ac similariations between the two, or by 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.

19
The internal structure of grains and their individuality was brought home to me through working with birthing. A particular song consisted of a rapid repeated sound having a "bubbling" quality. One might imagine from this that the individual sounds were internally articulated at a rate too fast to be heard. In fact, when slowed down to half its speed, each sound was found to be comprised of a rising scale passage followed by a brief portamento.

Larger sound events can often be described in terms of a onset or attack event and a continuation. The onset usually has the instantaneous 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 0.5 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 Beavague, 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 44).

If we try to make segments smaller than the grain-size, we will destroy the signal because the aperture (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-search by a factor of 2 we will in fact splice 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, but not the temporal structure, and we should achieve a time-stretching of the sound without changing its pitch (the harmonicization 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. (前所 example 3.1).

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 boundary between these time-frames (wavecycle, grain, continuation) are not, of course, completely clear cut and interacting perceptual ambiguities occur if we alter the parameters of a process so that it crosses these time-frame thresholds. In the simplest case, gradual 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-compressed into a structurally indecisive grain. (前所 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 (intonation), spectral contour or formant gliding (see Chapter 3) etc. The continuation may, however, be discontinuous as in noise sounds (noise-streams – such as called "R", and low contamination notes which are perceived partly as a rapid sequence of onsets) or segmented sounds (see below), or granular texture streams (e.g., murmur...
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. 130 crotchets per minute). In most music, bar-length is regular for long sections 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 replaces these together tail to head. In the simplest case sounds are selected in order, from the source, and replaced 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.,

<table>
<thead>
<tr>
<th>time</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>2.7</td>
</tr>
<tr>
<td>1.3</td>
<td>2.2</td>
</tr>
<tr>
<td>1.7</td>
<td>2.1</td>
</tr>
</tbody>
</table>

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.
The first significant object from a musical point of view is a shape made out of samples, and in particular a wavecycle (a single waveform 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 waveforms is repeated, a single wavecycle supplies no frequency information. Nor 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, timbre, spectrum etc) to the sound (for a fuller discussion of this point see On Sound Art).

This perceptual boundary is strikingly illustrated by the presence of waveart time-stretching. This process lengthens a sound by repeating each waveform, but not exactly the same thing. For more details see Appendix A).

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 rustle "R") by examining the loudness intensity of the sound. Intuitively we can say that the grains are the loud part of the signal, and the points between grains the quieter parts. If we set up an instrument which waits for the signal level to drop below a certain value (a threshold) and then count off the sound (grains), we should be able to separate the individual grains.

However, on reflection, we see that this process 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 process will clip the signal into discontinuous half wavecycles or smaller units (see Diagram 4) - not just as it was intended. What we must ask the instrument to do is search for a point in the signal where the signal stays below the (albeit) same 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 content, other characteristics (hard-edged, soft-edged), pitchiness/timbriness quality etc. The grain is characterized by a unique cluster of properties which we would be hard pressed to identify individually but which enables us to group it in a particular type e.g. unisoned "x", "y", "z", or "g".

Similarly, the spectral and pitch characteristics may not be easy to pin down, e.g. certain drums have a focused sound which we would expect from a pitched sound (they definitely don't have noise spectra as in a hi-hat), yet no atypical 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 in an inharmonic bell sound), or a rapidly periodizing 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.

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

CHANGE RINGING
A form of bell-ringing in which the order in which the bells are rung is determined 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 discriminately, in each block of 10 cycles per second (i.e. between 5 and 15, 15 and 25, 25 and 35 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 ending at the same time. Usually a set of pitches within a known reference set (e.g. the European tempered scale).

COMB-FILTER TRANSPOSITION............ 65
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).

COMB-FILTERING ............... 64
COMPRESSION

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. humming, sighing).

CONSTRUCTIVE DISTORTION

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

CONTAGION

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

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 Tapering for such temporal 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 as a single moment in time. But note that the spectral contour may itself evolve (change) over time.

CORRUGATION

CORCHETTE = (20) An indication of the speed at which musical events succeed one another. The duration unit, crochet, 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 of control, in as much a 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 to most commonly use in the time of writing (Aubert, 1994).

CUTTING

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 infrasonic 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 (used 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 many time-epochs (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-war does just this, see Appendix P1).

Note also that a very tiny fragment of the time-domain representation (a fragment shorter than a wavelength), although it may be meaningless 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 spread over the spectrum. Listening to it will we hear only a click (whichever the source).

Conversely, the frequency domain representation gives us more precise information about the frequency of the partial 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. There is a trade-off between temporal information and frequency information it is identical to the quantum mechanical principle of indeterminacy, where the time and energy (for 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 became aware of the different time-frames involved and their perceptual consequences. Stockhausen, in the article New Time Passes (published in Die Reihe) argued for the entry 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 "unly" mysticism which pervades much musical score analysis and commentary, but it does not tally with our 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.
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 common 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), through which 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 (e.g. performance gestures, some sonorants, represented in written scores. This abstracting process involves enormous compression in that it can only deal accurately with finite sets of definite phenomena and depends on existing musical assumptions about performance practice and instrument technology to supply the missing information (e.g. most of it). (See the discussion in On Sonic Art.) In contrast, the essential abstraction involved in digital recording leaves nothing to the imagination or fore-knowledge of its musicians, 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 analogs. 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 p.3).

However, perceptually we are more interested in the perceived properties of this sound. Is it pitched or messy for this? Can we perceive pitches within it? etc. In fact the feature underlying these perceived properties of a sonic sound is the disposition of its partials. Those may be conceived of as the simpler vibrations from which the actual waveform is constructed. It was Fourier (an Eighteenth Century French mathematician investigating heat distribution in solid objects) who realized that any particular wave shape (i.e., mathematically), any function, can be resolved by summing together elementary sinusoidal waves of the correct frequency and loudness (Appendix p.2). In acoustics these are known as the partials of the sound. (Appendix p.3).

If we have a regular repeating wave-pattem 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 even 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 single pitched sounds) what its pitch might be.

The spectral pattern which produces a single 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

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

As irreversible transformation of the waveform of a sound, changing its spectral quality (the brightness, sonorities etc. rather than the pitch or duration. Distortion also impiles the degradation of the sound (and irreversibility means that the original sound cannot be restored from the distorted version). Destructive distortion which preserves some-resonating points can be musically useful. See WAVESET.

DIPHOTE SYNTHESIS

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

DRONE

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

Ducking ........................................62

Mute of ensuring the prominence of a loud "sound" in a mix.

DURATION

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

DYNAMIC INTERPOLATION

The process whereby a sound gradually changes into a different sound(s) over the course of a single sonic event.
EDITING

The processes of cutting sounds into shorter segments and splicing together sounds or segments of sounds are called "splicing" or "looping." These techniques allow for creative manipulation of sound and are often used in various forms of production, from music to film. This chapter will explore the techniques and tools used for editing sound.

ENVELOPE

The loudness trajectory of a sound (the way the loudness varies over time) is often referred to as the envelope of the sound. Computer instruments which manipulate this loudness trajectory are usually called "envelope." 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 ...................................... 58,61
- ENVELOPE TRANSFORMATION .................................. 60

Musical transformations of the loudness trajectory of a sound are called "envelope." This chapter will explore how to use these techniques in different artistic contexts.

EXPANDING ......................................................... 60

- EXPANDING ......................................................... 60

Diagram 3

Sound source

microphone

amplifier

lateral disc cutting

Phonograph record

sound recorder

tape recorder

lateral disc playing

Phonograph

Sound

Loudspeaker

Ear

Diagram 3 continues on the next page...
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 consistency of instrument design and performance practice and the medium of the score, essentially a set of instructions for producing sounds 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 3.)

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 record 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 its essential part of the chain in the capture of the phenomenon of sound. The fundamental idea here is of the analogue. When we work on sound we in fact work on a spatial or electrical or mechanical 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 move at a consistent 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—analouges 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 disc) or the variations of magnetism along a tape (analogue tape recorder). By passing 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 recovered 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 g3.) The individual numbers are known as samples (not to be confused with chronicle of recorded sounds, recorded in digital samples, also often referred to as 'samples'), and each one represents the instantaneous (digital) value of the pressure in the original air pressure wave.

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.

PF
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 its...
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

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

GATING. 60

GOLDEN SECTION

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

AP:PB = PA:AP

this ratio is described as the Golden Section. It is approximately equal to 0.618

A = P = B

DIAGRAM 2

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

**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 wave.

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

---

**Grain**

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

**Grain Stream**

A sound consisting of a rapid sequence of similar events.

**Grain Time-Frame**

The typical duration of a grain.

**Granular Processes**

Musical processes which preserve the grains within a grain stream.

**Granular Reconstruction**

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 repositioned tail to head. Granular Reconstruction of output density 1 is terminage.

**Granular Reordering**

**Granular Reversal**

**Granular Synthesis**

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

**Granular Time Shrinking by Grain Deletion**

**Granular Time Shrinking by Grain Separation**

**Granular Time Stretching by Grain Duplication**

**Granular Time Stretching by Grain Separation**

**Granular Time Warping**

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

In traditional European practice, harmonic means pertaining to harmony where all the parts are multiples of some (audible) fundamental frequency. These parts are then known as harmonics. By this definition, 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 volume to pure, pure 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-size-frame processing 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 moment with.

HIGHLIGHT

Pitch in a sound which refers to a Harmonic Field or to European notions of Harmony. In contrast to pitch as a property of a spectrum.

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

The signal processing power of the Composer means that sound itself can now be manipulated. Like the chemicals, we can take apart what we are once the raw materials of music, reassemble them, or transform them into new and unknown forms 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 principles will still play strongly into the design of both musical tools and musical structures.

The precision of computer signal-processing means, furthermore, that previously evanescent and uncontrollable features of sounds may be analyzed, understood, transformed and transformed in rigidly definable ways. A minute usable 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 sound event can be pared away until only a few of the constituent partials remain, transforming something that was perhaps coarse and jumbled, into something archetypal (spectral tracing - see Chapter 3). We may exaggerate or constrain - in a precise manner - the energy (loudness) trajectory (or envelope) of a sound, enhancing or confining its gestural properties, and we can pass between these "states of being" of the sound with complex facility, tracing not an audible path of musical connections - a basis for musical form-building.

This shift in emphasis is as radical as it 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 "tonic matrix" 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 physical of sound, and how it can now be physically represented. Fundamentally sound is a perturbated wave travelling through the air. Just as we may observe the waves - swaying seaward from a stone thrown into still water, when we speak simple regular waves travelling through the air to the ears of listeners. And as with all water motion, it is the presence of disturbance which moves forward, rather than the water or air itself. Each pocket of air (or water) may be visualized as vibrating about its current position and passing on 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 when the air molecules move in mass in one direction. It also explains how sound can travel very much faster than the most locomotive 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) disperse the surface at right angles (up and down) to the direction of motion of that wave (forward). These are transparent to all lateral waves. In a sound wave, the air is alternatively compressed and rarified 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).
sounds corresponding to the various chapters in the book. In fact, sounds can be grouped into different classes, with flabby soundness, but these 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 metale 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 sonatas 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 close 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 pipes (ganstlando), and setting 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 accented through notation conventions. Most importantly however, apart from the field of perception, the overwhelming dominance of pitch as a defining parameter in music focuses interest on sound classes with relatively subtle 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, sit out those that were completely gold to make another instrument, and so on. The composer then becomes an expert in constructing maestros that in every pebble is of a definable color. As the twentieth century has progressed and the possibilities of conventional instruments have been explored to their limits we have learned to recognize various shades of gray and gray to make art architecture even more elaborate.

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

In a tonal sense, not only sounds of intemperate pitch (like unpitched percussion, defined portamenti, or inharmonic spectra) but those of unstable, or rapidly varying spectra (the gonging gas, the human speechstream) 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 teaching plans sound pitch (and duration) as their primary ordering principles, working with 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 combining 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

\[ \text{Inbetweening} \]

\[ \text{Indian Rag System} \]

\[ \text{Inharmonic} \]

\[ \text{Interpolation} \]

\[ \text{Inverse Fourier Transform} \]

\[ \text{Inverse Earplus Strong} \]

\[ \text{Sound Plucking} \]

\[ \text{Iteration} \]

\[ \text{Jitter} \]

They random fluctuations in apparently perceptually stable properties of a sound, much as pitch or loudness.
KARPLUS STRONG

An efficient algorithm for generating plucked-string sounds.

**Key**

A piece of music, in effect using the template scale (except where it is intentionally altered) 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.

**KLANGFARBEN-MELODIE**

Musical line where successive pictures are played by different groups of instruments. Literally, tone-colour melody.

---

**LIMITING**

60

**LINEAR PREDICTIVE CODING**

12

**LOOPING**

42

**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 for 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 sense the term loudness is used in the text whenever this will not cause any confusion. Diagrams usually refer to Amplitude.

**LPC**

12

---

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 purpose. 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 ideograms. "F5 flt" on a flute is related to "E-flats 9" on a clarinet, but these are relations between Ideals, or classes, of sounds. For every "F5 flt" on a flute there 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, timbral and vibrato control etc. are expected to be within perceptible limits defined by performance practice, but beyond this lies a whole less-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 area 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 distort 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; a particular occasion by an untrained speaker; the presence of a passing wolf in a particular forest or a particular vehicle in a road-tunnel; the entire extended solo in an all-time great improvisation which none 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 some 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 this 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 class of sounds that 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 groups (particularly onset-grains), pitch or pitch-band, pitch motion, spectral harmony-inharmonicity and its evolution, spectral contour and formants (see Chapter 3) and their evolution spectral stability and its evolution, and depressive, escalating and/or forced continuation (see Chapter 4). All of the same time. In dividing up the various properties of sound, we don't wish to imply that there are different classes of
We already know that sound composition presents us with many more 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 as we say "this is sound composition as we know it" in the same sense that we can do this with traditional instrumental composition.

***** 5 *****

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 variations).

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, sustain), together with other instrument specific data (which synthesizer 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.

MIX-SHUFFLING

MIXING

MIXING SCORE

A set of instructions dictating 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 notation 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 even that we can visualize in our own spatial experience) from the fine dimensions of Einstein’s space–time, to the infinite number of dimensions in Hilbert space.

MULTI-SOURCE BRASSAGE

***** X *****

NOISE

Sound having no perceptible pitches) 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 consistent "X" or "A". Other sounds (especially those recorded directly from the natural environment) may contain degrees 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
credentials as an "original" artifice. 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 hostile is performing on a new instrument which is composed of the complete system acoustic-instrument-plus--electronic-network. Any new instrument taken to master. Hence there is a change that a piece for electronically processed acoustic instrument will fall short of our musical expectations because no matter how well 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 practices, 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 repertory for these new instruments develops.

The second paradigm is that of pure electro-acoustic (studio) composition. Such compositions may be regarded as suffering from the drawback that it is not a "belief your very eyes," and interpretable medium. In this respect it has the drawbacks, but also the advantages, that film has via-a-via theater. What film lacks in the way of irreplaceable archetypality(?) it makes up for in the closely observed detail of location and specifically captured human sensation. Similarly, studio composition can deal with the unisonness of sound 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 speak of "archetype" I mean that the same elements that are contained in an example of a class of possible sounds which all satisfy the restrictions placed on being a flute callid "F". 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 are used as the basis of compositional processes which depend on these uniquenesses. The power of the computer in both sound 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 the result the studio composer must take on many of the functions of the performative as sound production and interpretation.

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 (whenever 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 run the program and subsequently heard the result. If I didn't like this result, I had to modify the values in the table and ran the program again, and so on. It is clearly simple and more efficient, when first exploiting any time-varying process and its effect on a particular sound, to move a slider, turn a knob, how a string, blow down a tube (etc) and hear the result of the physical action as I make it. I can then adjust my physical actions to satisfy my mental experience - I can explore intuitively, without going through any conscious explanatory control process.

OCTAVE
A sound whose pitch is an octave higher than a second sound, usually has twice the frequency of that second sound. Two pitches in the same octave in equal tempered music, are regarded in some sense as the "same" pitch, or as belonging to the same 'pitch-class'.

OCTAVE-STACKING

ONSET SYNCHRONISATION

PARAMETER
Any property of a sound or a sequence of sounds which can be musically organized. Parameter often also implies the measurability of that property.

PARTIAL

The sinusoidal elements which define the spectrum of a sound.

PARTIAL TRACKING

PERMUTATION
Specific rearrangement of the elements, or the properties of the elements (e.g. loudness), of a sequence of musical events.

PHASE

PHASE INVERSION

The wave form 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 inverting the wave from the first zero at which it reaches 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.

8

126
PHASE VOCODER .......................... 11

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 one definition is more subtle.

PHRASE
Element of musical structure consisting of a sequence of sound events and, in traditional practice, usually having for a few bars.

PHYSICALITY
The physical nature of the apparent source of the sound. NOT the physical source of the real source. In the apparent source hard or soft, rigid or flexible, granular of or-a-piece etc.

PITCH ................................. 4
A property of instrumental and vocal sounds organized in most traditional musical practices. Pitch arises from a regular arrangement of particles in the spectrum. All particles 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. Human hear pitch between approximately 16 cycles per second and 4000 cycles per second. Below 16 cps, the sound breaks up into a gase-stream. Above 4000 cycles we may still be aware of intonatia (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) parts of a sound on a different sound.

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 thereby placed on the computer to choose an appropriate set of configurations for a particular purpose. The "instrument" is no longer a definable (if nebulous) closed universe but a ground swell of possibilities out of which the sonic composer must define 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 decisions 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-uncontrollable 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 score (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 duality of real-time processing has on compositional practice apart from speeding up some of the more mundane, technical involved.

Although the three traditional roles-performer, instrument-builder and composer are being blurred by the new technological developments, they provide useful points 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, in which the composer provides electronic extensions to a traditional instrumental performance. This approach is intrinsically "real-time". The advantages of this paradigm are those of "breath" (the theatre versus the cinema, the work is interrupted "before your very eyes") and of immiency dependent on each performer's interpretation of the work. This approach fits well into a traditional musical way of thinking.

It's drawbacks are not perhaps immediately obvious. But the composer who specifies a network of electronic processing devices around a traditional instrumental performance must recognize that he/she is in fact creating a new and different instrument for the instrumentation (or instrumentation-like) device to play and is partly adopting the role of an instrument builder with its own very different responsibilities. In the acoustic cultural atmosphere of the late Twentieth Century, the temptation for anyone labelled "composer" is to build a new electronic instrument for every piece, to establish
On the other hand, we also have explicit knowledge of, e.g. acceptable Harmonic progressions within a given style or the spectral content of a particular voice formant (see Chapter 3). This is knowledge that we know we have 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 means of explicitness is enhanced because it results in a document (the score) which can often be given a rational analysis by the composer or by a music score-analyst. Some composers (particularly the presently deceased "romantic" composers), may use the musical score nearly as a script for their intuitive material outpourings. Others may occasionally do the same but in a cultural atmosphere where explicit rational decision-making is more highly prized, will claim, post hoc, to have worked it all out in an entirely explicit and rational way.

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

The absent and misguided rationalist viewpoint 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 body 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 it desirable: in a medium where everything is possible, it is an essential part of the compositional process. In this case a synthesis is taking place between formally and performatively activities and as his activity of composer and performer begins 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 organization 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 file-improvising performer relies on an intuitive creative process (restrained by the increased sound/pitch limitations of a musical source, e.g. a reissued piano, a metal shake). To generate both the moment-to-moment articulation of events and the ongoing time structure at all levels.

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

PORTAMENTO

Sliding of pitch. Often incorrectly referred to as glissando, but there is an important distinction. On a furred or keyed instrument like a piano, we may slide fingers from a high pitch to a low pitch, but the fingers 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 notion 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-FOCUSED TRANSFORMATION

Some musical processes will be radically altered on a source sound that the perceived "ground-sound" is more dependent on the process of transformation than on the source itself. The same process applied to two different sources will produce very similar goal-sounds. The perceived result of the procedure is governed more by the arcana of the process of transformation, rather than by the particular nature of the source-sounds employed.

**** q ****

0 ______________________________________ 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 musical unit e.g. a grid of demi-semiquavers. All events must then fall on some multiple of demi-semiquaver divisions of the beat. The actual time quantisation will then be determined by the tempo (the number of crotchetts per second). The quantisation grid provides a time reference-frame. On keyed or furred 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). 128
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 resonate 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.

****** g ******

SAMPLER
A piece of hardware, or a software package on a computer, which 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 members used to record the shape of the waveform itself, which are properly known as samples. (See SAMPLING.)

SAMPLING ................................. 1

Sound is digitally recorded by samplinng 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 reproduced, e.g. between 22,050 and 48,000 samples per second. At a sampling rate of 48,000 samples per second, the highest resolvable frequency is 24,000 cycles per second.

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 mode of acoustic knowledge. An arbitrary, number-crunching program will produce an arbitrary result — like the old stage of trying to write a play by letting a chimpanzee type away at a typewriter in the hope that a newsworthy revelation will emerge.

So science is not maudlin? We may embark on signal-processing procedures which will appear banal to the scientifically sophisticated, procedures that give relatively unpredictable results, or that are heavily dependent on the shape properties of the particular signals to which they are applied. The question we must ask as musicians whenever we use are these procedures scientifically valid, or even predictable, but rather, do they produce any useful results at all on some types of sound materials?

There is also a word of caution for the composer reader. Music is a twentieth century Western art music has been dogged by an obsession with complication. This has arisen partly from the permutations procedures of late serialism and also from an intellectually biased dislike of crude information theory with musical communication — more precisely means more information, means more musical "power". This obsession with quantity, or information overload, arises partly from the breakdown of commonness in the substance of musical meaning, to the end, the meaning of music is delusive as ever it was in an age which demands everything be quantifiable and measurable, a model which stresses the quantity of information, or complication of an artifact, seems fatally 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 any 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 worlds of time and effort to do this. This does not mean however that I will, or even intend to, use every conceivable option the new instrument offers. It is just as with the traditional acoustic instruments, the task is to use it to achieve the end, to play it, well. In sound composition, this means I need 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 creates an immediate result which is monitored immediately by the ear 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 we do it. In music, intuitive knowledge is most strongly associated with musical performance.
The practical implications of this are immense for the composer. In the past I might spend many days working from an original sound source, substituting it to many processes before arriving at a satisfactory result. As a result of this process (as well as a safeguard against losing my hard-won final product) I would keep copies notates and copies of many of the intermediate sounds, to make recreation (or variations) 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 given 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 in the process of changing them as transformations or sound-transformations. The tools which effect changes will be described as acoustic 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 conceived as a form of activity which can be further developed and intensified in any community. We can be sure that this will be a highly active and enriching process. However, we should be aware that this is a work in progress and that the nature of this community can be quite different from the "new-fangled" or essentially "incorrect" or "inappropriate" seen in some musical or aesthetically unsound activity. Nevertheless, scientific readers will be more familiar with the terms signal, signal processing and the mathematical description of such a signal. In some cases, scientists share these goals. Precise analysis of sounds, extraction of time-varying information in the frequency domain (see Appendix II), sound clarification or sound reduction are all of great importance to the sonic composition. But beyond this, the question of how one creates music in, this process aesthetically interesting? Does the sound resulting from this process reflect perceptually and in a musically useful manner to the sound we use with? What are we searching for? Is this way to transform sound material to give meaning sounds which are clearly close relates of the source, but also clearly different? Ideally we require a way to "measure" or to these degrees of difference allowing us to anticipate the space of sound possibilities in a more structured and meaningful way.

Beyond this, there are no intrinsic restrictions on what we do. In particular, the goal of the process is not in motion may not be known or even (with complex signals) easily available beforehand. In fact, as musicians, we do not need to "know" completely what we are doing (12). 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 suit a particular need. A musician, however, is more likely to require an extremely large (see-here, time-variable, Q-variable, time-variable; Appendix II) 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 acoustic transformation of the original source. What this means in practice may only emerge in the course of the exploration.

SCORE

The notation of a system of music from which a performance of the work is recreated.

SEMISTONE

The smallest interval between consecutive pitches on a modern European keyboard or fisted instrument. Musical scales are defined as a set of tones equally spaced between semistones. Aromatic minor scales also contain a three semistone step.

SEQUENCE GENERATION

SEQUENCES

Consecutive sets of consecutive sounds having distinctly different spectral properties (e.g. speech, or melodic phrases on keyed instruments the species 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 is used as the basis for the organization of pitches in a piece. The general idea of indeterminacy was also extended to sequences of durations, of loudness, etc. The more general notion of pervading a given set of elements has been widely used (e.g. systems music).

SERIALISM

See SERIAL COMPOSITION

SHAWM

A medieval wind instrument with a single reed musician.

SHEEPHERD 'S HORN

Sounds (or sequences of sounds) constructed so that they rise in triads while their pitch falls (or vice versa).

SINE WAVE

The elementary oscillations in terms of which all other regular oscillations, vibrations or waveforms can be described. The oscillation of a simple (idealized) pendulum is described by a sine wave.

SINEWAVE

Having the shape of a sine-wave.
SOUND REVERSING .......................... 43
SOUND SHREDDING .......................... 41
SOUND-FLICKERING .......................... 72

A musical transformation that imposes a picked-string-like attack on a sound.

SOURCE-FOCUSED TRANSFORMATION

A musical transformation whose outcome depends strongly on the source sound. Defined in contrast to PROCESS-FOCUSED TRANSFORMATION.

SPECTRAL ARPEGGATION .......................... 24
SPECTRAL BLURRING .......................... 26
SPECTRAL BRIGHTNESS .......................... 18

A measure of where energy is focused in the spectrum. If some of upper periods 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

The world of sound-coproduction has been hampered by being cast in the role of a poor relation to more traditional musical practice. In particular the vest body of analytical and critical writings in the musicology of Western art music is strongly oriented to the study of musical sounds (sounds) rather than to a discipline of acute auditory awareness in itself. Sound composition requires the development of both new listening and awareness skills for the composer and, I would argue, a new analytical and critical discipline founded in the study of the sonic experience itself, rather than its representation in a text or score. 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 medicalized "text writing" on the critical-analytical disciplines which have evolved in music. Both the scientific method and technological advances in medical science have had to struggle against the passive authority of texts declaring eternal truths and values inherited from the scientific method and to technological advances.

I don't wish here to decry the idea that there may be "universal truths" about human behavior 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 preservation, 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 deriving 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 observation of sounds themselves. (Scientific method may be surprised to hear that this must 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 claims 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 perfect texts of our time 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, in particular to create new and unheard sonic experiences. Such texts are potent but, at the same time, justiciable. They do not radiate some mystical authority to which we must bow, 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.
WHAT THIS BOOK IS NOT ABOUT

This book is not about the arts of computers or particular programming packages. However, most of the programs described were available on the Composer Desktop Project (CDP) System at the time of writing. The CDP was developed as a composer 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 the course, most of them will run in real time environment.

My concern here, however, is to uncover the nuural possibilities and remains 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 MIDIS controller). Many composers are either forced into this approach, or do not see beyond it, because cheaply available technology is based on the instrument concept of music. As its simplest such an approach is no more than traditional non-oriented composition for electronic instruments, particularly where the MIDI interface confuses 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 prelearned pitch/Harmonic structure (though we may choose to adopt this approach in particular circumstamce) 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 hardly 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 emphasize 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 mentally perceived, on my direct response to these perceptions and on what can be technically, acoustically or mathematically described.

SPECTRAL TIME-SHINING 30-31
SPECTRAL TIME-STRETCHING 30-31
SPECTRAL TIME-WARPING 30-31
SPECTRAL TRACE-AND-BLUR 27
SPECTRAL TRACING 25
SPECTRAL UNDEFORMATION 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.

SPELING 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

SRLT

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 one 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.

STERE0

Sound emanating from two channels (e.g., two loudspeakers) 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 occurring 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 randomization, 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 recognizably such, but now it is possible, through a process of careful analysis and subtle transformation, to recreate a recorded sound in a changed from which however sound as convincingly `natural' as the recording of the original sound.

TAPE-ACCELERATION ...................... 36
TAPE-SPEED VARIATION ...................... 36
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.

\[
2^1, 2^2, 2^3, 2^4, 2^5, 2^6, \ldots
\]

and the frequency ratios between members of this series are known as `pens' intervals. Some pen interval ratios are...

<table>
<thead>
<tr>
<th>Octave</th>
<th>400</th>
<th>405</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>505</td>
<td>510</td>
</tr>
<tr>
<td>700</td>
<td>705</td>
<td>710</td>
</tr>
</tbody>
</table>

There is no common smaller interval from which all these `pens' intervals can be constructed. Hence the surviving 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 pen 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 crochet is as long as two quavers. The speed of the whole will be indicated by a tempo marking e.g.

crochet = 120

which means there are to be 120 crochets in one minute.

WHAT THIS BOOK IS ABOUT

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

1. Any sound whatever 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 intimate 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 Varèse was imagining a then unavailable music which had the same degree of control over sonic substance as musicians had traditionally exercised over melody, harmony and rhythm. This concern grew directly out of the sophisticated development of orchestration in the last Nineteenth Century and its intimate 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 emerging and increasing potential of the technology of sound recording which made this dream accessible.

The exploration of the now sounds made available by recording technology was begun by Pierre Schaeffer and the G.R.M. in Paris in the early 1950s. Initially harnessed by unsophisticated tools (in the early days, editing between tape dos - here the transformations - like tape speed variation, editing and uncutting - available with magnetic tap) masterpieces of this new medium began to emerge and an approach to musical composition round 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 can 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 mass 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 so, to date, no universal tradition of large scale forms-building (through these newly available sound-relationships) has established itself as a norm.
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, amplified (e.g., robed) 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

Organization of sound elements in terms of (temporal) density and field properties.

TEXTURE CONTROL ......................... 68-69
TEXTURE GENERATION ...................... 68-69
TEXTURE OF GROUPS ....................... 68

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 OF MOTIFS ...................... 69

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 preordained arrangement (e.g. some combination of sequence, time-placement, pitch-relationship, formant sequence, etc).

TEXTURE OF ORNAMENTS .................. 69

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

Devised 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 set 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, wavelet time-stretching, harmoniser, brassing.

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.

TOWAL MUSIC
Astatic organisation around keys and the progression between different keys. In contrast, serial music avoids indicating the dominance of any particular key or pitch.

TONES
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 sounders".

TRIMODO .................... 66
Cyclical modulations of loudness between c. 4 and 20 cycles per second.

TRANPOSITION
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.

CONTENTS

Chapter 0 Introduction .................. 1
Chapter 1 The Nature of Sound ............ 11
Chapter 2 Pitch ..................... 25
Chapter 3 Spectrum ................... 35
Chapter 4 Onset ...................... 44
Chapter 5 Continuation ................ 48
Chapter 6 Grain Streams ............... 55
Chapter 7 Sequences .................. 59
Chapter 8 Texture .................... 66
Chapter 9 Time ...................... 72
Chapter 10 Energy ................... 81
Chapter 11 Time-stretching ............. 86
Chapter 12 Interpolation ............... 96
Chapter 13 New Approaches to Form .... 103
Appendix 1 ........................ 112
Brief Bibliography .................. 139
TRITONE

The musical interval of 6 semitones. In the European tempered scale, a frequency ratio of the square root of 2, to 1.

UNDULATION

cylical change in the value of a musical property, like the rising and falling of pitch in vibrato, or of loudness in tremolo.

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

Cyclical undulations of pitch between c. 4 and 20 cycles per second.

VOCODING

WAVELENGTH

WAVESET

WAVESET DISTORTION

Generally refers to any process which intentionally alters the waveshapes in a sound e.g. wavelet inversion, omission, reversal, shifting, skewing, 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).
WAVESET ENVELOPING ............... 53
WAVESET HARMONIC DISTORTION .... 52
WAVESET INTERLEAVING ............. 54
WAVESET INVERSION ................. 51
WAVESET OMISSION ................. 51
WAVESET REVERSAL .................. 51
WAVESET SHAKING ................... 51
WAVESET SQUEEZE ................... 51
WAVESET STATIONARY ............... 52
WAVESET STRETCHING ............... 55
WAVESET TRANSFER .................. 54
WAVESET TRANSPOSITION .......... 51
WEDGEING .......................... 60

WHITE-OUT
A musical process in which a texture becomes so dense that all spectral detail is lost, producing a noisy sound.

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 basic 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.

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 Chassan Buczok, in 1993. Some clarifications and supplementary material was added after discussions with Miller Puckette, Zak Stefan, Stefan Bilharz and Philippe Despain at IRCAM. Moreover, the whole of any inaccuracies or omissions in the exposition rests entirely with me.
The text of the book was originally written entirely in longhand and I am indebted to Wendy McGeeery for typing the personal metaphysical hieroglyphics into computer files. Help with the final text layout was provided by Tony Mynt at 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 like to thank my fellow contributors to this long running project; in particular, Tim Endrich, the main driving force of the CDP, to whom we owe the survival, expansion and development of this congenerious venture, against all odds, and whose persistent probing and questioning has led to clear instrument descriptions & musician-friendly documentation.

Richard Ornstein and Andrew Bentley, the other composer founder members of, and core instrument contributors to the project: David Malham who devised the hardware basis of the CDP and taught electrics, and continues to give support; Martin Atwood, 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 anachronistically, if not yet elegantly!); Rajnish Fuchsman of the University of Keele, who has been principally responsible for developing the various graphic interfaces to the system; to Michael Clarke, Nick Lawson, Rob Waring, Richard Debrauwere in the many students at the Universities of York, Keele, and Birmingham and to individual users everywhere, who have supported, used and helped sustain and develop this resource.

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

137
ZERO-CROSSING

The waveform of a sound continually rises above the center line and then falls below it. The point where it crosses the center line is a zero-crossing.

ZERO-CUTTING ........................ 40

ZIGZAGGING ........................... 43
TRITONE
The musical interval of 6 semitones. In the European tempered scale, a frequency ratio of the square root of 2, to 1.

***** x *****

UNDULATION
cyclical change in the value of a musical property, like the rising and falling of pitch in vibrato, or of loudness in tremolo.

***** x *****

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 30 cycles per second.
VOCODING .......................... 34

***** x *****

WAVELENGTH ......................... 3.4
WAVESET ............................ 50
WAVESET DISTORTION .................. 52
Generally refers to any process which incrementally alters the waveforms in a sound e.g. wavelet inversion, omission, reversal, shifting, attenuation, 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).

136
BRIEF BIBLIOGRAPHY

DIE REBKE
Journal published by serial composers associated with the Darmstadt Summer School in the 1960s.

GESANG DER JUNGLINGE
Electronasonic work by Karlheinz Stockhausen, using boy’s voice and electronics. Recording available.

GRUPPEN
Large scale work for 3 orchestras surrounding the audience, by Karlheinz Stockhausen. Recording available.

ON SONIC ART

PYTHOPRAKTIA
Work by Iannis Xenakis for orchestra, exploiting masses of portamenti and using statistical formulae to control textures.

RED BIRD
45 minute electro-acoustic work by Trevor Wishart (1973–77). Recording available.

VOX 3
Work for 4 amplified voices by Trevor Wishart, from 1987, in which the voices sing in polyphony, coordinated by the use of independent but computer-synchronised click-tracks. Recording available.

VOX 5
Electro-acoustic work by Trevor Wishart, from 1986, in which a human voice is transformed into the sounds of many other recognizable sounds. Recordings available.