The Development of a Cognitive Framework for the Analysis of Acousmatic Music

David John Godfrey Hirst

Submitted in partial fulfilment of the requirements of the degree of Doctor of Philosophy (by creative work and dissertation)

May 2006

Faculty of Music
The University of Melbourne

Produced on archival quality paper
Abstract

This PhD in Music Composition is in the form of a written dissertation plus creative work. The creative work is a series of electroacoustic music compositions on Audio CD, submitted as Appendix 1.

The analysis of acousmatic music has traditionally been very difficult since there is no score to "freeze" the music in time. Analysis relies heavily on the act of concentrated listening. Since aural perception is so crucial to the analysis of acousmatic music, this dissertation poses the questions: Can a framework for the analysis of acousmatic music be derived from cognition theories, research on the auditory perception of everyday environmental sounds, and studies into the perception of Western tonal music? If so, what are the framework's attributes?

From experimental data documented in the relevant literature, this dissertation draws together the constituents of a cognitive framework. The framework has been called the Segregation, Integration, Assimilation and Meaning (SIAM) framework for the analysis of acousmatic music.

The dissertation reports on the practical application of the SIAM framework through a detailed analysis of the work Wind Chimes, by Denis Smalley. The analytical method uses a combination of signal analysis and critical listening. Observations were recorded in text and a pictorial forms. A multimedia "interactive study score" was created to provide a dynamic visual representation along with synchronized playback of the recorded work. Use of the interactive study score is documented in Appendix 2, and a functional interactive study score is provided on the included CD-ROM.

Analysis of the subsequent data has led to some of the following conclusions regarding the organisation of Denis Smalley's Wind Chimes: Smalley extensively manipulates internal spectral structures. He pushes traditional tonal music concepts higher in the frequency spectrum, creates compound sound events, and makes extensive use of the attack-resonance model. The primary binding force is proximity in frequency, and the form of the work is sectional, episodic, and much like an expression of the attack-resonance model on a macro scale. Interpretation of the
analytical data has also resulted in the emergence of certain syntactic traits: compound sound events; stratification of the frequency spectrum; pitch centricity and harmonic fields; organisation in time; semantic and syntactic progression.

The dissertation asserts that the SIAM framework developed for the analysis of acousmatic music has been effective in that: It provides a systematic frame of reference for a thorough examination of acousmatic musical works by addressing the detail through the adoption of the SIAM procedure, and by addressing the "big picture" through the use of the listener interpretation-attention notion. The framework has resulted in a number of new insights into the particular work Wind Chimes. However, to be fully tested and refined into a robust methodology, the framework needs to be applied to other acousmatic works.

Finally the dissertation discusses some implications for future research, including: The use of models of the human auditory system; 3-D representations that could be developed in connection with the auditory modelling techniques; and experiments to investigate the role of grouping and hierarchy formation in the experience of acousmatic music.
Declaration

This is to certify that:

i. The thesis comprises only my original work towards the PhD except where indicated in the Preface,

ii. Due acknowledgement has been made in the text to all other material used,

iii. The thesis is less than 100,000 words in length, exclusive of tables, maps, bibliographies and appendices.

Signed:..................................................................................(David Hirst)
Preface

This PhD in Music Composition is in the form of a written dissertation plus creative
work. The creative work is a series of electroacoustic music compositions on Audio
CD, submitted as Appendix 1. Also included is a data CD-ROM which contains an
interactive study score created by the author. Notes for its operation are given in
Appendix 2.

In accordance with one of the options for a PhD in Music Composition at the
University of Melbourne, the dissertation’s research topic is largely independent of the
creative work in this submission, although there are some mutual influences which are
referred to in the Appendix 1 notes.

Composition of the musical work called Mon Dieu was actually begun prior to the
commencement of candidature, but it has not been submitted for any other degree at
any other university. The submitted version was completed in 2002.
Acknowledgements

I would like to give acknowledgement and my heart-felt thanks to the following: Professor Brenton Broadstock for his guidance and assistance; Dr Neil McLachlan for eternal hours spent in enthusiastic encouragement, debate and review; Professor Denis Smalley for permission to reproduce *Wind Chimes* in the interactive study score and for making himself available for such generous conversation and correspondence; the University of Melbourne for material assistance; staff of the University for help with candidature matters; and to Raelene Reece, Caitlin Hirst and James Hirst for their continual support throughout this journey.
Table of Contents

Abstract i
Declaration iii
Preface iv
Acknowledgements v
Table of Contents vi
List of tables, figures and illustrations xi

CHAPTER 1 : INTRODUCTION 1

CHAPTER 2 : AUDITORY SCENES AND TIMBRE ORGANISATION 5

2.1 Bregman’s Auditory Scene Analysis 5
  2.1.1 Auditory Scene Analysis Summary 6
  2.1.2 Primitive Auditory Scene Analysis 7
  2.1.3 Sequential Integration 8
  2.1.4 Sequential Integration Factors 8
  2.1.5 Spectral Integration 10
  2.1.6 Spectral Integration Factors 10
  2.1.7 Schema-based Segregation 12
  2.1.8 Auditory Scene Analysis and Music 13

2.2 Smalley’s Timbre Theories 14
  2.2.1 Defining Timbre 15
  2.2.2 Smalley’s Theories of Timbral Organisation 16
    Timbre and Source 16
    Source-cause Texture 17
    Registration 19
    Existence 20
    Coherence 20
    Discourse Stability and Variability 21
    Transformational discourse 22
    Typological discourse 22
    Discourse Summary 23
    The Timbre of Pitch – the Pitch of Timbre 23
    Smalley’s Concluding Remarks on Timbral Organisation 24
  2.2.3 Developing an Analysis Methodology Based on Smalley’s Theories 25

2.3 Conclusion 26

CHAPTER 3 : EVENT STRUCTURE PROCESSING 27

3.1 Analytical problems within acousmatic music 27

3.2 Some frameworks for the analysis of acousmatic music 28

3.3 Developing a cognitive framework 30
  3.3.1 Why a cognitive model? 30
  3.3.2 Bigand’s cognitive model for tonal music 31
3.3.3 Tonal Hierarchies
3.3.4 Lexicon of schemas
3.3.5 Event Structure Processing
   Reading the acoustic surface
   Auditory image formation
   Time-span segmentation
   Time-span reduction (Establishing event hierarchies)
   Prolongation reduction (Perception of patterns of musical tension and relaxation)
   Symbolization processes

3.4 A cognitive framework for acousmatic music
   3.4.1 Basic Assumptions
   3.4.2 Event Structure Processing in Acousmatic Music
      Reading the acoustic surface
      Auditory image formation
      Auditory scene recognition
      Time-span segregation
      Time-span reduction
      Prolongation Reduction

3.5 Conclusion

CHAPTER 4 : SIGNS AND SYMBOLS, SYNTAX AND SEMANTICS

4.1 Introduction

4.2 Signs, symbols and emotions in tonal music
   4.2.1 Index
   4.2.2 Icon
   4.2.3 Symbol
   4.2.4 Emotion : A cognitive theory
      Arousal
      Meaning
      Arousal, complexity, and stylistic change

4.3 Recognition of signs and symbols in acousmatic music
   4.3.1 Recognition of sound sources and events
   4.3.2 Identification of environmental sounds
   4.3.3 Research on the perception of environmental sounds
   4.3.3 Gygi’s Definitions and Experiments
      Similarity studies
      Identification of filtered environmental sounds
      Identification of environmental sounds from event-modulated noise
      Acoustic factors in EMN identification
   4.3.4 Factors in the identification of environmental sounds

4.4 Syntactic and semantic factors
   4.4.1 Background to the work of Howard and Ballas
   4.4.2 Experiment one – Syntactic factors with simple tones
   4.4.2 Experiment two – Syntactic factors with environmental sounds
   4.4.3 Experiment three – Syntactic and semantic factors with environmental sounds
   4.4.4 Contribution to identification by acoustic, semantic, syntactic and other factors
      Experiment One: Identification time and causal uncertainty
      Acoustic factors in identification time and accuracy
      Experiment Two: Ecological frequency study
      Experiment Three: Perceptual and cognitive judgements
      Summary of the Ballas experimental findings
CHAPTER 6 : APPLYING THE ANALYTICAL METHODOLOGY 141

6.1 Aims 141

6.2 Previous Methods of Representation 141

6.3 Practical Analytical Methodology 143
  6.3.1 Digital Encoding 143
  6.3.2 Segmentation and Section Formation 143
  6.3.3 Initial Analytical Process 144

6.4 Design of the Interactive Study Score 145

6.5 Results 147
  6.5.1 Section 1 148
    Observations 148
    Discussion 150
  6.5.2 Section 2 151
    Observations 151
    Discussion 152
  6.5.3 Section 3 154
    Observations 154
    Discussion 156
  6.5.4 Section 4 157
    Observations 157
    Discussion 159
  6.5.5 Section 5 161
    Observations 161
    Discussion 166
  6.5.6 Section 6 170
    Observations 170
    Discussion 172
  6.5.7 Section 7 174
    Observations 174
    Discussion 177
  6.5.8 Section 8 179
    Observations 179
    Discussion 181
  6.5.9 Section 9 183
    Observations 183
    Discussion 184

6.6 Analysis of the Whole Work 184
  6.6.1 Methodology 186
  6.6.2 Progression within the categories 188
    Texture 188
    Frequency Organisation 190
    Organisation in Time 191
    Semantic and Syntactic Progression 192

6.7 Summary 195
  6.7.1 Compound sound events 195
  6.7.2 Stratification of the frequency spectrum 195
  6.7.3 Pitch centricity and harmonic fields 196
  6.7.4 Organisation in time 197
  6.7.5 Semantic and syntactic progression 198

6.8 Conclusions 198
CHAPTER 7 : SUMMARY AND CONCLUSION 201

7.1 Framework for Analysis 201
  7.1.1 The SIAM Framework 201

7.2 Application of the Framework 204
  7.2.1 Methods and Tools 204
  7.2.2 Notational Forms 207
    Interactive Study Score 208
    Annotated Study Score 208
    Short Score 209
  7.2.3 Insights 211
    Compound sound events 213
    Stratification of the frequency spectrum 213
    Pitch centricity and harmonic fields 214
    Organisation in time 214
    Semantic and syntactic progression 215
  7.2.4 Effectiveness of the Framework 216
  7.2.5 Implications for Future Research – or “Unanswered Questions” 217
    Hierarchy in Acousmatic Music 217
    Further Experiments 220
    Patterning, Progression and Prolongation 221
    Implications of Future Research Resulting from the Current Study 222

Bibliography 224

Appendix 1 – Composition Folio 231
  La vie naturelle 231
  Maison 231
  Artificier artificiel 232
  Travail 233
  Mon Dieu 233

Appendix 2 – Supplementary CD-ROM 235
  Interactive Study Score 235
  Instructions for Use 235
List of tables, figures and illustrations

TABLE 3-1: TIMBRAL HIERARCHY WEIGHTINGS ................................................................. 55
TABLE 3-2: COMPARISON OF TONAL AND ACOUSMATIC DISCOURSES ....................... 59
TABLE 4-1: HEVNER’S EMOTIONAL INTERPRETATIONS OF MUSICAL FEATURES .......... 64
TABLE 4-2: WEDIN’S ANALYSIS OF LISTENERS’ ADJECTIVES ASCRIBED TO MUSICAL FEATURES .......................................................................................................................... 65
TABLE 4-3: IMPLICATIONS OF MICRO-TEMPORAL AND MACRO-TEMPORAL PROPERTIES ................................................................................................................................. 70
TABLE 4-4: STRUCTURAL FEATURES OF SPEECH, MUSIC AND ENVIRONMENTAL SOUNDS ................................................................................................................................. 78
TABLE 4-5: GROUPING OF SOUNDS INTO FIVE CLASSES ACCORDING TO EVENT TYPE. 83
TABLE 4-6: DIFFERENCES BETWEEN PRIMITIVE AND SCHEMA-BASED SCENE ANALYSIS ................................................................................................................................. 105
TABLE 4-7: NARMOUR’S PARAMETRIC SCALE ................................................................. 120
TABLE 5-1: THE SIAM FRAMEWORK .................................................................................. 131

FIGURE 2.1: OVERVIEW OF BREGMAN’S AUDITORY SCENE ANALYSIS ............................. 7
FIGURE 2.2: PRIMITIVE AUDITORY SCENE ANALYSIS ....................................................... 8
FIGURE 2.3: SCHEMA-BASED SEGREGATION ..................................................................... 13
FIGURE 2.4: CONCEPTS AND RELATIONSHIPS DEVELOPED BY SMALLEY (1994) .................. 25
FIGURE 3.1 BIGAND’S MODEL OF EVENT STRUCTURE PROCESSING .................................. 35
FIGURE 3.2 MCADAM’S STAGES OF PROCESSING IN AUDITORY RECOGNITION ..................... 46
FIGURE 3.3: SOUND EVENT PROCESSING ........................................................................... 60
FIGURE 4.1: PLOT OF LISTENER PREFERENCE VS MUSICAL COMPLEXITY ......................... 68
FIGURE 4.2: THE 2-D MDS SOLUTION FOR GYGI’S “ACOUSTIC SIMILARITY” EXPERIMENT ................................................................................................................................. 81
FIGURE 4.3: STATE-TRANSITION DIAGRAM USED BY HOWARD AND BALLAS .................... 88
FIGURE 5.1: LISTENER INTERPRETATIONS OF ACOUSMATIC MUSIC .................................. 138
FIGURE 6.1: SCREEN SHOT OF THE WIND CHIMES “INTERACTIVE STUDY SCORE” ............... 146
FIGURE 6.2: WIND CHIMES STUDY SCORE 0’-2’ ................................................................ 148
FIGURE 6.3: WIND CHIMES STUDY SCORE 2’-4’ ............................................................... 154
FIGURE 6.4: WIND CHIMES STUDY SCORE 4’-6’ ............................................................... 157
FIGURE 6.5: WIND CHIMES STUDY SCORE 6’-8’ ............................................................... 169
FIGURE 6.6: WIND CHIMES STUDY SCORE 8’-10’ ............................................................ 174
FIGURE 6.7: WIND CHIMES STUDY SCORE 10’-12’ ......................................................... 178
FIGURE 6.8: WIND CHIMES STUDY SCORE 12’-14’ ......................................................... 182
FIGURE 6.9: WIND CHIMES STUDY SCORE 14’-END ....................................................... 184
FIGURE 6.10: TABULATION OF CATEGORICAL FEATURES FOR SECTIONS OF WIND CHIMES – PAGE 1: 0’ – 4’ ...................................................................................... 187
FIGURE 6.11: TABULATION OF CATEGORICAL FEATURES FOR SECTIONS OF WIND CHIMES – PAGE 2: 4’ – 8’ ...................................................................................... 187
FIGURE 6.12: TABULATION OF CATEGORICAL FEATURES FOR SECTIONS OF WIND CHIMES – PAGE 3: 8’ – 12’ ................................................................. 188
FIGURE 6.13: TABULATION OF CATEGORICAL FEATURES FOR SECTIONS OF WIND CHIMES – PAGE 4: 12’ - END ................................................................. 188
Chapter 1 : Introduction

This thesis is concerned with a genre of music called *acousmatic music*. Acousmatic music is a style of electroacoustic music that uses recordings of everyday environmental sounds, recordings of unprocessed instrumental sounds, processed environmental sounds, processed instrumental sounds, and synthesized sounds to create musical works that are generally written to a “fixed medium” such as an audio CD or tape.

Some writers and researchers in this field can tend to use the terms *electroacoustic music* and *acousmatic music* in an equivalent fashion, but electroacoustic music can have a broader interpretation too. Throughout this thesis, the term acousmatic music will be used, except where another author is quoted.

Luke Windsor has provided this useful definition of acousmatic music:

The term acousmatic music is here partly interchangeable with the term *musique concrète* as originally intended by Schaeffer and reiterated by Michel Chion (1993). Acousmatic music is music that is recorded and then diffused without combination with live electronics or live performers; it exists only on tape (whether analog or digital) or as a fixed set of instructions to a computer. The term *acousmatic* is preferred to *concrète* as it emphasises the way in which the ‘real’ acoustic source is assumed to be hidden from the audience. Acousmatic music does not exclude the use of synthesis or sound processing; but these processes must be employed in the making of a fixed artefact that is then diffused, rather than employed during performance. (Windsor, 1995: 11)

The analysis of acousmatic music has traditionally been difficult since, unlike Western tonal music, there is no standard score format. Listening becomes the primary focus in the analysis of acousmatic music. If listening is the focus, then the study of auditory perception and cognition may provide a firm basis for the establishment of a rigorous methodology for the analysis of acousmatic music. The primary research question therefore becomes:

- Can a procedure for the analysis of acousmatic music be derived from the theories of auditory perception and cognition, and if so, what are its attributes?
The sensory source stimulus is acousmatic music, but there are scant theories on the perception and cognition of acousmatic music as there have been virtually no psychological experimental studies on its perception. This observation gives rise to several secondary research questions:

- Can the results of studies into the perception of Western tonal music be adapted to contribute to a framework for the analysis of acousmatic music?
- Can the results of studies into the auditory perception of everyday environmental sounds be adapted to contribute to a framework for the analysis of acousmatic music?

A final set of supplementary research questions derive from the application of a cognitive framework to the analysis of acousmatic music. They are questions related to methodology:

- What methods and tools must be adopted to apply the cognitive framework to the practical analysis of repertoire acousmatic works?
- What notational forms must be derived to communicate any analytical findings in a written medium?
- Is the framework an effective approach to the analysis of acousmatic music?
- What new insights result from the application of the constructed framework and methodology?
- What are the implications for subsequent research?

The scope and style of this study occupies the territory between the disciplines of music psychology and musicology. For example, no experimental studies on the perception of acousmatic music have been undertaken in this work, but rather the music psychology and general cognitive psychology literature has been explored in order to develop a theoretic framework for the analysis of acousmatic musical works, based on previously-documented aspects of the perception of sound. Part of the framework aims to test a hypothesis regarding the adaptability of the results from
perceptual studies of other “musics” and the perception of everyday environmental sounds.

The methodology, documented here, initially takes the theoretical framework developed and combines it with the analysis of data resulting from signal processing (spectral analysis) along with critical listening of the acousmatic musical work being studied – Denis Smalley’s *Wind Chimes*.

The chapters that follow document the journey undertaken to answer the research questions posed above. The true introduction to the subject of this thesis begins with Chapter 2 which outlines two different viewpoints. Chapter 2 begins with a presentation of the ideas of a cognitive psychologist who specialises in auditory perception. An alternative perspective is that of an acousmatic music composer, whose writings are a reflection on the processes and forces that are at play in the realisation and perception of his works. At first glance, the two viewpoints may seem to be quite different theories, but closer scrutiny reveals that there are many overlapping concepts. Their ideas are a fundamental preamble to the analysis of acoumatic music. Not only do these two writers provide fundamental perspectives vital to this study, but they also lend a vocabulary for the discussion of the chimeric experience called acousmatic music.

In Chapter 3, the focus is on the microscopic, moment to moment, experience of listening, and a model from the literature on the perception of Western tonal music is outlined – *Event Structure Processing*.

Chapter 4 examines the macro-view with reference to relevant research on signs, symbols, syntax, semantics and their interpretation. Research on environmental sound perception and musical structuring forms the basis for this chapter.

Chapter 5 sets out to unify the concepts presented in the previous two chapters into a single methodological framework that can be applied to the analysis of acousmatic music – the SIAM framework.

Chapter 6 applies the framework to an extensive analysis of Denis Smalley’s *Wind Chimes*. 
In the final chapter there is a discussion in relation to the original research questions and some suggestions with respect to areas that require further investigation.
Chapter 2: Auditory Scenes and Timbre Organisation

This chapter presents two perspectives on the perception of environmental sounds and music - the views of a scientist and the ideas of a music practitioner. While it is tempting to portray them as complete contrasts, we will find that there are a number of commonalities between their concepts. The theories of Bregman (1999) and Smalley (1994) are singled out for special attention, not only because both have been very influential in their respective fields, but also because their unique insights draw upon vocabularies and concepts that will prove useful in developing a cognitive framework for the analysis of acousmatic music.

2.1 Bregman’s Auditory Scene Analysis

Undoubtedly one of the most substantial bodies of work on the perception of ordinary everyday sounds is the work carried out by Bregman (1999) on auditory scene analysis. Bregman argues that the study of audition has been the poor cousin of the study of vision. Prior to 1965, textbooks discussed basic auditory qualities like loudness and pitch, and the physical properties of the sound that give rise to the perceptual quality we experience.

Rather than discussing auditory perception as a series of fragmented phenomena or theories, Bregman tries to solve the problem of what he calls “auditory scene analysis” in order to create a more unified theory.

Let me clarify what I mean by auditory scene analysis. The best way to begin is to ask what perception is for. ... Perception is the process of using information provided by our senses to form mental representations of the world around us. In using the word representations, we are implying a two-part system: One part forms the representations and another uses them to do such things as calculate appropriate plans and actions. (Bregman, 1999: 3)

Bregman discusses some other difficulties associated with auditory scene analysis. He says that: "... in vision, you can describe the problem of scene analysis in terms of the correct grouping of regions. But what about the sense of hearing? What are the basic parts that must be grouped to make a sound?" (Bregman, 1999:6)
To help the discussion Bregman introduces the visual representation of a single sound via the spectrogram. The corollary is that we may use this blueprint to recognise the sound at a later date. The complication is that the world doesn't consist of pristine individual sounds, but rather there is a mixture of many sounds at any one point in time. Bregman gives the example of a spectrogram depicting the spoken word “shoe” mixed with other sounds.

The recogniser would have to solve the following problems: How many sources have created the mixture? Is a particular discontinuity in the picture a change in one sound or an interruption by a second one? Should two dark regions, one above the other in the picture (in other words occurring at the same time), be grouped as a single sound with a complex timbre or separated to represent two simultaneous sounds with simple timbres? (Bregman, 1999:9)

In order to recognise a picture of blocks visually, regions must be grouped appropriately. This can be done by taking a set of crayons and colouring in surfaces. In a similar manner, the regions on a spectrogram that come from the same source could be “coloured in” and this is how Bregman describes the auditory scene analysis problem.

### 2.1.1 Auditory Scene Analysis Summary

Bregman describes the problem of auditory scene analysis as a problem of deconstruction of simultaneous and sequential sounds.

Although we need to build separate mental descriptions of the different sound-producing events in our environment, the pattern of acoustic energy that is received by our ears is a mixture of the effects of the different events. It appears that our auditory systems solve the problem in two ways, by the use of primitive processes of auditory grouping and by governing the listening process by schemas that incorporate our knowledge of familiar sounds. (Bregman, 1999: 641)

Figure 2.1 provides a conceptual overview of Bregman’s auditory scene analysis. At the top of the diagram, auditory scene analysis branches into primitive auditory scene analysis and schema-based analysis. Primitive auditory scene analysis dominates the discussion in Bregman’s book.
2.1.2 Primitive Auditory Scene Analysis

The primitive process breaks down incoming acoustic energy into separate frequency bands. The problem then becomes one of how to group the frequencies that may be emanating from the one sound source, in order to identify each separate sound source. Bregman suggests that there are two forms of grouping. Simultaneous grouping he calls “spectral integration”, and temporal grouping he describes as a “sequential integration”. Figure 2.2 displays these main concepts associated with primitive auditory scene analysis along with their contributing factors.
2.1.3 Sequential Integration

Sequential integration is observed when frequencies that occur one after the other are grouped together in a chain to form an auditory stream. Bregman quotes this example of auditory streaming:

The simplest case is a rapid repeated alternation of a high and a low tone. If the alternation is fast enough, listeners will not experience a single stream of tones alternating in pitch, but will perceive two streams of tones, one consisting of repetitions of the lower tone and the other consisting of repetitions of the higher tone. (Bregman, 1999: 642)

2.1.4 Sequential Integration Factors

*Frequency proximity and temporal proximity:* Bregman postulates that there are two primary competing factors at work in the segregation of frequencies into sequential streams. The first factor, primitive segregation, is influenced by the rate at which the total sequence unfolds, and by the frequency separation between the interleaved tones:

As the frequency separation increases, the sequence must be slowed down if the listener is to be able to experience all tones as a part of a single, coherent stream of sound. (Bregman, 1999: 643)
The second main factor at work in the segregation and integration of frequencies into separate sequential streams, is schema-based segregation that involves attention. If listeners are trying to focus their attention on the tones of one of the streams, the effects are different:

The frequency separation of the high from the low tone need only exceed some small amount (a few semitones in the case of two alternating tones) before the target sequence can be followed by attention. (Bregman, 1999: 643)

Here we can see the first mention of the attitude of the listener and the effects this attitude might have.

Bregman notes that stream formation principles are analogous to the Gestalt principle of grouping by proximity, both proximity in frequency and temporal proximity.

Spatial origin: Spatial origin is another factor influencing sequential integration. Where sounds emanate from different possible locations in space, we tend to group sounds coming from the same point. On the other hand, listeners can also do quite well at segregating sounds coming from a single point in space.

Brightness: Most of the research work on sequential integration has concentrated on tests using pure tones, but timbre can also influence the grouping into streams. Bregman points to the difficulty that, unlike pitch and loudness, timbre is multi-dimensional. Bregman singles out “brightness” as one dimension of timbre that can provide a basis for similarity between tones. Brightness is the mean frequency of all frequencies weighted by their own loudness.

Spectral shape: Spectral shape could also be used as a basis for similarity checks. Peaks at the same place in the spectrum of two sounds could imply that they have passed through the same resonator, whereas equal proportions between the harmonics and fundamental frequency of two tones would imply that the properties of the original vibrating medium of each are similar. What little evidence that exists suggests that both tests of spectral similarity are used to group successive tones.

Hysteresis: Bregman also comments on a certain “hysteresis” or persistence in perception of around four seconds.
Once some interpretation of a number of sonic sources has occurred it does not disappear instantly just because some source has not been heard for a second or two. (Bregman, 1999: 648)

Interestingly, a sudden change in acoustic properties can reset this hysteresis mechanism faster than silence can.

It is important to note that there is competition as well as collaboration between the various factors influencing sequential integration:

If a number of factors all favour the grouping of sounds in the same way, the grouping will be very strong and the sounds will always be heard as parts of the same stream. (Bregman, 1999: 651)

Returning to the idea of listener attention, Bregman puts forward a theory of “attention as a filter” that is tuned to certain properties and can change its settings. All sounds must pass to be part of the same act of attention. Attention settings can’t be changed rapidly so sudden changes can be missed.

Bregman also notes that Gestalt grouping is like pre-attentive grouping:

It sees the effects of similarity, temporal proximity, and continuity as innate principles that determine grouping. The idea of competitive forces is also part of the theory. (Bregman, 1999: 654)

2.1.5 Spectral Integration

When we experience a mixture of sounds, our auditory system must decompose the simultaneous frequency components, and then recompose them into separate groups that appear to come from separate sources. This process of simultaneous grouping Bregman calls “spectral integration”.

2.1.6 Spectral Integration Factors

Bregman puts forward a number of strategies for the grouping of simultaneous frequencies.

Old-plus-new heuristic: A complex spectrum may contain a simple spectrum that was just heard earlier. This theory postulates that the part of the complex spectrum
matching the earlier tone is a continuation of it and therefore separates out from the complex mixture. This is really just a variation of the sequential grouping principle.

**Frequency relations:** wide frequency separation implies the frequencies are from separate sources. Intense partials are easier to segregate since they are less prone to the phenomenon of masking.

**Harmonicity:** Many vibrating physical bodies produce harmonic spectra, with partials that are whole number multiples of the fundamental. Finding a certain number of fundamental frequencies that can account for all partials present implies that there are the same number of environmental sources. The detection of these fundamental frequencies also allows us to associate a pitch with each group, thus leading to the perception of multiple, simultaneous pitches.

**Common fate:** “The Gestalt psychologists discovered that when different parts of the perceptual field were changing in the same way at the same time, they tended to be grouped together and seen to be changing as a group because of their common fate.” (Bregman, 1999: 657)

Correlated changes in frequency of partials represent one manifestation of common fate. For example, micro-modulation in the pitch of the human voice can occur, even when you think you are holding a steady pitch. Slow modulation occurrences can be voluntary such as a singer using vibrato, or a speaker raising pitch at the end of a sentence. Synchronised amplitude change is another version of common fate. Micro-modulations and macro-modulations can affect amplitudes too, for example partials starting and stopping at the same time will tend to be from the same source.

**Spatial origin:** There is some evidence to suggest that separate bands of frequencies can be assigned to different spatial origins, and this can be used to reinforce other integration factors.

**Other factors:** Bregman notes that textural features resulting from phenomena such as torn paper, or walking on crunchy snow, may contribute to partitioning, but there has been little research in this area. Rhythm may play a role too, but is this a primitive process or a schema-driven one?
2.1.7 Schema-based Segregation

Bregman contrasts primitive auditory scene analysis with schema-based scene analysis. Schemas are units of mental control containing knowledge of the environment:

Each schema incorporates information about one particular regularity in the environment. Regularity can occur at different levels of size and spans of time. So in our knowledge of language, for example, we would have a schema for “a”, one for the word “apple”, one for the grammatical structure of a passive sentence, one for the pattern of give and take in a conversation, and so on. (Bregman, 1999: 666)

Primitive auditory scene analysis is concerned with sound properties whereas schema-based analysis is concerned with pattern recognition. Schemas involve attention (to listen for something) which requires the subjective experience of effort. The number of things that can be attended to simultaneously is quite limited. Attentive listeners always have some conception of what they are listening for.

Schemas become active when patterns are detected. Patterns tend to extend over time, and when part of the evidence of a pattern is detected, the associated schema is activated and it prepares for the perception of the rest of the pattern. Bregman calls this preparation schema-governed attention. A schema can also be activated by an associated schema, so reading the word “apple” may prepare for reading “fruit”.

Bregman provides a good example of the schema-based selection of sense data in the phenomenon of phonemic restoration:

… in the phonemic restoration of a speech sound that has been masked by a sudden loud noise (we hear the speech continue through the noise). Apparently we select certain frequency components out of the noise and hear them as if they were the missing speech sounds. (Bregman, 1999: 667)

Figure 2.3 highlights the main concepts associated with schema-based segregation.
Here we can summarise schema-based segregation and contrast it with primitive scene analysis. A schema is generally associated with a particular concept and a single regularity or pattern in the environment. The patterns can be different levels of size or different spans of time. Schemas apply to classes of sounds and can be activated by other schemas.

Schemas are learned, whereas primitive processes appear to be present in new-born babies and are therefore said to be innate. A schema selects from the evidence in an asymmetrical way, whereas primitive processes partition sensory evidence equally (and symmetrically). Primitive processes utilise short-term properties of sensory data, while schema-based selection involves looking at relations over a longer time span.

2.1.8 Auditory Scene Analysis and Music

Bregman’s book concentrates on primitive auditory scene analysis rather than schema-based organisation. He applies the principles he has developed for primitive auditory scene analysis to the role of primitive organisation in music.
Obvious parallels are that melody formation may involve some form of sequential integration (temporal grouping), while timbre and chordal harmony are the result of spectral integration (simultaneous grouping).

The consequences of the application of Bregman’s auditory scene analysis to the analysis of acousmatic music will be explored in subsequent chapters, but this section will close with an interesting quote from Bregman concerning hierarchy and music:

There is a second type of segregation between the different parts of the piece of music itself. This organisation has to be strong in some ways and weak in others. The segregation has to be strong enough for the listener to perceive each line of melody with its own distinct timbre. At the same time it must not be so strong as to keep us from perceiving the musical relations between the parts. The best solution would be to organise the music into a hierarchical form (parts within larger parts). Think of the way in which the fingers on a hand are perceived: distinct and yet united to create a larger form. (Bregman, 1999: 675)

Bregman continues by noting his belief that perception is structured in this sort of hierarchical way but cannot say whether such hierarchical structures are created by primitive or schema-based processes, although he suspects the latter.

How will this strong versus weak segmentation interplay be manifest within the acousmatic works to be analysed? This is a question that will be taken up again in later chapters. For the moment, a quite different approach to the discussion of perception, timbre, environmental sounds and music will be considered, and this is the approach taken by Denis Smalley, summarised in the next section.

### 2.2 Smalley’s Timbre Theories

Denis Smalley has written and composed extensively within the field of acousmatic music. In his writings he has developed new concepts and theories to describe acousmatic music. Probably the best known concept is his spectromorphology - the temporal unfolding and shaping of sound spectra (Smalley, 1986).

Rather than survey all of Smalley’s written works, this chapter concentrates on one of his seminal articles called *Defining Timbre – Refining Timbre* (Smalley, 1994). A substantial part of what follows is an attempt to summarise and make comment on the Smalley article. It is also an attempt to demystify many of Smalley’s concepts since
they are shrouded in, sometimes difficult, terminology which is much of his own
invention. Although the jargon is difficult, it does offer the prospect of facilitating
dialogue concerning some very hard-to-grasp acousmatic music concepts. Having
identified Smalley’s salient theoretic concepts, the rest of this chapter aims to codify
the relationships between the concepts in a way that may prove useful in analysing
such a unique genre of music. The processes of identifying entities and establishing
functional relationships are just as relevant to musical analysis as they are to musical
theoretic concepts.

Organising timbre can be a primary driving force in electroacoustic music, but timbre
has had an infamous reputation for being an ephemeral term. Smalley begins his
article with the following caution:

(Timbre) is one of those subjects where the more you read and the more you have
hands on compositional experience the more you know, but in the process you
become less able to grasp its essence. (Smalley 1994:35)

2.2.1 Defining Timbre

The difficulties in defining timbre are even highlighted by the contradictions in
naming timbre. While the French term *timbre* identifies the object that actually creates
the sound, the German term *Klangfarben*, or sound colour, is more abstract and
detaches the sound from any source.

Like Grey (1975), Erickson (1975) highlights the multi-dimensional nature of timbre,
and its subjective and objective variations. He talks of the subjective constancy of
timbre and its use as a carrier of other musical information (e.g. melody) and of
timbres as objects. These come together in *Klangfarbenmelodie* as contrast and
continuity.

Denis Smalley presents four definitions from four different perspectives:

The American National Standards Institute:

...that attribute of auditory sensation in terms of which a listener can judge that two
sounds similarly presented and having the same loudness and pitch are dissimilar.
(Smalley, 1994:36)
The instrumental composer:

Timbre is an extension of harmony, or vice versa. The composer uses spectral analysis as a basis for conceptualising the relationship between pitch and sound qualities, and attempts to negotiate fluent border crossings between the two. (Smalley, 1994: 36)

The researcher:

Through research publications and through electroacoustic compositional experience we have become very aware of the multiple variables which determine timbral identity. And we have also become concerned to differentiate what is acoustically present in sounds from what is psycho-acoustically pertinent. (Smalley, 1994: 36)

Everyone:

The everyday language of qualitative description is accessible to everyone. It is closely allied to the "matter" of sound. Terms like bright/dull, compact/spread, hollow, dense... (Smalley, 1994: 36)

2.2.2 Smalley’s Theories of Timbral Organisation

In this section, a short summary of the concepts that Smalley has developed to explain the organisation of timbre in acousmatic music is presented.

Timbre and Source

Smalley highlights the importance of the sound source when discussing timbre. He provides a preliminary definition that takes the sound source into account (adapted from Michel Chion):

Timbre is a general, sonic physiognomy through which we identify sounds as emanating from a source, whether the source be actual, inferred or imagined. (Smalley, 1994:36)

Smalley defines the term source bonding to encapsulate:

The natural tendency to relate sounds to supposed sources and causes, and to relate sounds to each other because they appear to have shared or associated origins. (Smalley, 1994:37)
Smalley then acknowledges the socio-cultural contribution the listener makes in the "adventure of bonding play" when listening to electroacoustic music, which he regards as a perceptual activity.

Smalley explains that source bonding is extrinsic in that it refers to sounding experiences outside the work itself. He then defines the extrinsic matrix which, in addition to external sounding experiences, includes links to a wide range of real and imagined non-sounding phenomena too. (Smalley, 1994:37)

Smalley doesn’t expand on the notion. Instead he turns to the notion that timbre is concerned with the temporal unfolding and shaping of sound spectra or spectromorphology. Motion, growth and energy are associated with spectromorphology and therefore have a sonic reality, but they can also be interpreted metaphorically and symbolically.

We can see that in addition to the physical unfolding of the temporal shape of the sound (amplitude, spectrum) Smalley has added the listener, metaphor, symbology, and perceived source. Smalley attempts to define this complex combination in an abstract and succinct way. His definition of timbre has become:

A general, sonic physiognomy whose spectromorphological ensemble permits the attribution of an identity. (Smalley, 1994: 38)

**Source-cause Texture**

In a section called "source-cause texture", Smalley invents a lot of his own jargon which can be difficult to interpret. It is worth persisting, however, as many of his ideas strike at the heart of the function of timbre in acousmatic music.

He begins this section by looking at the identity of instruments, exposed through our experience of what he calls source-cause levels.

*Imminent level* – is the ongoing, intrinsic musical context where we encounter the instrument (eg. the violin). Associated with the imminent level is registration - the articulation of note objects and their chaining in phrases over a continuum of registers.
*Cumulative level* - which includes our previous experiences of violin sources in the hands of other violinist-causes who articulate the same music and other genres and styles.

*Extended level* - extends the source-cause base to include the immediate family of stringed sources. (eg. viola, 'cello, double bass)

*Dispersed level* - spreads over the widest possible range of source-causes to include all bowed and plucked (string) instruments and the (string) instruments of other cultures.

The four levels together he calls source-cause texture. We may accept this scheme for instrumental music, but how can we apply his notion of source-cause texture to acousmatic music? Smalley answers by noting that in acousmatic music we do not find such a definable hierarchical basis for establishing the source-cause aspect of timbral identity. To illustrate, he turns to the concrete example of a "water" source-cause. "In a musical work, water, like any sounding source, can exist on both the imminent and cumulative levels." (Smalley, 1994:39) That is, the water sounds can function within this musical work and other musical works, but it is much harder to identify families of water sounds or extended families of water sounds (as they function within musical works). So although the extended and dispersed levels cannot be identified in acousmatic music, Smalley goes outside the confines of the music:

...it can be extended beyond the cumulative level by referring outside the musical works to the extrinsic matrix. (Smalley, 1994: 39)

That is, to both the sounding and non-sounding inferred phenomena:

... the sounding area of the extrinsic matrix can provide a substitute for the extended and dispersed levels of source-cause texture ... (Smalley, 1994: 39)

So in our example of water, there are many experiences of water sounds beyond the cumulative level, in nature and culture. But what about non-sounding inferences?

When the non-sounding area of the extrinsic matrix is entered, we are no-longer on secure common ground because source bonding is no longer operable: we cannot identify real sources and causes. (Smalley, 1994: 39)
Smalley does, however, go on to highlight the importance of such non-sounding inferences and contends that it is the spectromorphological attributes and ideas that evoke non-sounding substitutes for (instrumental) extended and dispersed levels. There is some resonance with the ideas of the ecological psychologists here.

While all of this seems very ephemeral, it may be the very essence of being for acousmatic music.

I believe that such ideas, intangible as they might be, are a means of articulating that necessary, shared, higher-level cultural basis for a music with non-existent source-causes. (Smalley, 1994: 39)

Smalley begins to expand on the idea that spectromorphological attributes can evoke non-sounding real and imagined phenomena by first considering instrumental gesture.

... behind the causality of instrumental gesture lies both a broader experience of the physical gesture and its proprioceptive tensions, and a deeper, psychological experience of gesture. In instrumental music human-bonded source-cause texture represents these primal levels of gesture found in the extrinsic matrix. In electroacoustic music where source-cause links are severed, access to any deeper, primal, tensile level is not mediated by source-cause texture. That is what makes such types of acousmatic music difficult for many to grasp. In a certain physical sense there is nothing to grasp - source-cause texture has evaporated. (Smalley, 1994: 39)

Smalley is desperately searching for a "Something to hold on to factor" (Landy, 1994) when considering the most abstract aspects of acousmatic music.

Finally, he turns to the spectromorphology of the musical work in the search for the identity of timbre.

Registration

Smalley states that timbral identity in acousmatic music is heavily reliant on the immediate experience of the work itself and, moreover, the ready identification of the functional elements in the work.

If timbral identity in electroacoustic music is so heavily reliant on the single imminent level then registration in the work is key. (Smalley, 1994: 40)
I interpret his concept of registration in music as being the recognition and connection of events or phenomena in the music. While this has been systematically codified by musical theorists in instrumental music, it is problematic in acousmatic music where there is no consistent low-level note object.

Registration then becomes concerned with variable spectromorphological attributes. Smalley notes that this could become so general we could lose our way, so we need to discriminate the incidental from the functional. We need to perform a functional analysis of acousmatic music.

We have to decide which spectromorphological attributes matter, and we have to discover this anew in the imminent level of each acousmatic work. A pair of related variables underpins our attempts at determining identities:

1. The coherence and strength of spectromorphological identity.
2. The duration needed to establish existence and expose registration. (Smalley, 1994:41)

**Existence**

In the search for the identity of fundamental elements within a work (eg. separate sound events), there is no standard duration for establishing existence.

There are certain types of spectromorphology whose existence can only be established after a certain evolution time because the completion or partial completion of a pattern is integral to identity. (Smalley, 1994: 41)

In addition, the application of a continuous transformation to a sonic entity that is ever-changing incrementally can become what Smalley calls a registration generator. Change becomes fundamental to existence and we apprehend identity as a consequence of change.

**Coherence**

Smalley introduces coherence by stating:

The notions of existence and registration imply a certain coherence, otherwise identity would not be feasible. (Smalley, 1994: 41)
To my mind, the process of searching for coherence is the same as the process of creating connections within a work. Smalley, however, considers coherence in a much more subtle way, on both the micro and macro levels. He uses it to establish identity. Along the way he creates the notions of integration and disintegration.

Coherence in the case of instruments is usually associated with spectral fusion, which in turn is associated with harmonicity. ...Since fusion is so often closely aligned with harmonicity, and since I find it too rigid a word, I prefer the term "integration". That also allows us to talk of an integration-disintegration continuum. (Smalley, 1994: 41-42)

Smalley also highlights the problem of discriminating between integration and disintegration: “When is a timbre a timbre and when is it a collection of timbres? When does the physiognomy crack?” (Smalley, 1994: 42)

His solution to creating connections between identities in electroacoustic music is to introduce the concept of discourse:

Playing with integration and disintegration is at the very heart of electroacoustic musical discourse, a discourse which becomes spectromorphology itself once the timbre complex is spilt open, a discourse where the notion of timbre can at one moment perhaps be grasped, but at the next it evaporates. (Smalley, 1994: 42)

**Discourse Stability and Variability**

In acousmatic music, a timbre may be a discrete object, easily separated from its context or it may be a continuity intertwined with other continuities and not so easily separated. To help identify this phenomenon, Smalley introduces the concept of timbral level, which concerns the relationship between two continua:

- **Duration continuum**: short-term entity --- longer term evolution
- **Separability continuum**: discrete object --- continuing context

Smalley notes that timbral level in traditional note-based music is quite simple:

The note is the lowest level and is articulated by an instrumental source. Form develops from note articulations. In electroacoustic music continuing contexts resist and deny low-level segmentation. Thus once timbral level ceases to be clearcut we cannot separate timbre and discourse: timbral attributes become woven into the spectromorphological fabric. (Smalley, 1994: 43)
Now that Smalley has established some bases for determining identity he discusses discourse in more detail. He asserts that discourse is about maintaining and developing some of the established identities.

**Transformational discourse**

The first type of discourse Smalley defines as transformational discourse “where an identity is transformed while retaining significant vestiges of its roots”. (Smalley, 1994: 43) He then gives some examples of prevalent techniques used for digital transformations.

Smalley makes some important points in passing regarding memorability and contiguity.

... electroacoustic music which avoids pitch phenomena and strong source-cause references will be most frequently concerned with contiguous relationships. … Non-contiguous discourse can only be highly developed within the precise and detailed memorability of culturally imbedded pitch and rhythm systems. (Smalley, 1994: 44)

The above quote contains important implications for the analysis of Smalley’s works. For example, where pitch, rhythm and strong source-cause references are not used then we should find contiguous relationships and continuous transformations.

**Typological discourse**

Smalley's second type of discourse is typological discourse.

Identities are recognised as sharing timbral qualities but are not regarded as being descendants of the same imminent identity – they do not possess a common identity base. Typological discourse is associative. (Smalley, 1994: 44)

Smalley invents the term *generic timbres* to describe larger groupings of timbres. He gives two examples of generic timbres: noise and inharmonicity.

To transformational and typological discourses Smalley adds a third type: *source-cause discourse* – concerned with the bonding play of specific or inferred sounding identities.
Discourse Summary

Smalley points out that these various discourses are not mutually exclusive, and “tipping the scales either way can be a question of listening choice and listening experience – listening experience is variable and subjective.” (Smalley, 1994: 45)

Taking the listener into account, Smalley ends up with six interactive types of electroacoustic discourse. The first three are those mentioned already: source-cause discourse; transformational discourse; typological discourse. The next set of discourses are concerned with relations among identities. They are:

- Behavioural discourse – the changing states of identities’ cohabitation/conflict and dominance/subordination;
- Motion discourse – the relations of types of motion and growth and their directional tendencies;
- Tensile discourse – how the previous five discourses together create formal tensions.

(Smalley, 1994: 46)

The Timbre of Pitch – the Pitch of Timbre

Smalley also writes of the pitch of timbre and the timbre of pitch. This section is a very important and very revealing discussion in regard to Smalley’s own works. Pitch is an important defining variable in Smalley’s works, but its manipulation is not overt and its use is hard to grasp. He introduces the notion that pitch and timbre can cohabit the same space.

Once tonality and intervallic pitch are no longer regarded as the predominant carriers of musical messages, pitch and timbre can cohabit in a spectromorphological music where the ear has opportunities for shifting in and out of pitch values. (Smalley, 1994: 40)

Surprisingly Smalley postulates that: “Pitch is even present when not perceived.” (Smalley, 1994: 40). This statement may provide a telling insight when his own works are examined, but how can a non-perceived pitch exist, when by definition pitch is perceived phenomenon? Smalley elaborates:
Perhaps it is resting, hidden deep in a spectromorphology, awaiting possible attention, a moment when, for example, the context might change so that perceptual focus becomes directed towards what was a sleeping attribute. (Smalley, 1994: 40)

It would seem that, for Smalley, pitch perception can be dependent upon context in acousmatic music. Smalley also observes that timbre runs the risk of becoming meaningless since it is often described as the sound attribute that is not pitch. So if pitch and timbre are under threat of dissolution in acousmatic music, then what is the glue that binds this music together?

**Smalley’s Concluding Remarks on Timbral Organisation**

In his concluding remarks regarding timbre, Smalley rightly confesses that he has not been concerned with the acoustic nature of timbre, nor with its psychoacoustic properties, but with its apprehension, identity and functions in musical contexts from the listener’s point of view. For Smalley, timbre has become the timbre complex which has been spilt open to allow its constituents to seep through the musical discourse.

Composing with timbre, composing within timbre, means confronting and enjoying its dissolution. This can only be really pursued in an acousmatic electroacoustic music. (Smalley, 1994: 47)

Re-assuring us that traditional modes of working and listening are still valid, Smalley concludes:

There is no reason why the traditional notion of timbre should fade away. A notion of musical timbre will always exist alongside its dissolved attributes. In keeping with this ambivalence I can summarise this discourse in six words: Timbre is dead. Long live timbre. (Smalley, 1994: 47)

Figure 2.4 illustrates some of the main concepts, actions and relationships developed by Smalley.

---

The Development of a Cognitive Framework for the Analysis of Acousmatic Music  
24
2.2.3 Developing an Analysis Methodology Based on Smalley’s Theories

What if we were to develop a protocol for the analysis of acousmatic music, like the music that is composed by Denis Smalley, based on his own theoretical constructs? What would that protocol look like?

An inside-out approach would look at an unfolding of the moment-to-moment details of the music itself, and could begin by examining the imminent and/or cumulative levels via the registration of spectromorphologies.

The steps may involve:

- Establishing the identities by looking at coherence, integration and disintegration.
- Establishing functional relationships between the identities.
- Identifying transformational, typological, and source-cause discourses.
- Identifying behavioural and motion discourses.
- Discussing the tensile ebb and flow.

An examination of the extended and dispersed levels would reveal more of the extrinsic nature of the work. The process would involve connecting the internal sound
world of the work to external sonic manifestations and to the more ephemeral, non-
sounding, conceptual implications.

We can see from the above description that Smalley's auditory scenes are very
complex, and the way his “scenes” are connected with each other within his works is
crucial to how listeners interpret the works.

2.3 Conclusion

Although they are quite different approaches, the writings of Smalley and Bregman
provide us with the beginnings of a vocabulary to describe some aspects of acousmatic
music. Interestingly, the writings of the composer and the psychologist do have some
similarities and overlaps. Smalley even uses the term “integration” which is so
fundamental to the segregation-integration ideas found in Bregman.

In addition, they both leave some unanswered questions. What are the relative
contributions of the bottom-up approach (primitive processes) and the top-down
approach (schema-based)? How can we apply these concepts to a specific musical
example?

So, although Bregman and Smalley have provided us with some of the vocabulary and
conceptual framework for use in the analysis of acousmatic music, we need more
detail, and a concrete methodology to follow. What other research can contribute to a
perception-based methodology for the analysis of acousmatic music?

The next couple of chapters will seek some answers to these questions.
Chapter 3: Event Structure Processing

One of the innovations of Twentieth Century music was the ability to record segments of sound and include them (treated or untreated) in musical works. The inclusion of recorded sound segments, which were often environmental sounds, brought its own set of compositional problems for the composer, as well as interpretational challenges for the listener.

This chapter aims to shed some light on the analytical challenges in acousmatic music by focusing on the listening process as an expression of real-time analysis. Some clues are then sought from the field of cognitive psychology as to how one could structure the analysis of acousmatic music in a more systematic way.

3.1 Analytical problems within acousmatic music

Many of the difficulties experienced in the analysis of acousmatic music were recorded in an article by Marco Stroppa (1984). His major criticisms can be summarised under two headings.

The problem of the score: For music that exists solely on tape, there is limited effective visual representation. There are traditionally two types of what may be called a “score” - a list of operational data used to generate the piece, or a sketch of the musical effects obtained.

The former suffers in that it is only useful in regenerating (reproducing) the work itself. The latter is guilty of being “always rather crude and approximate, particularly in comparison with the complexity and perfection of traditional notation. Analysts in search of compositional method, or something more profound than a simple observation of contrast or similarity will find themselves up against a brick wall.” (Stroppa, 1984: 177-178)

The problem of perception: Stroppa (1984: 179) criticises the “Let's get rid of the written text, and think more about what happens to our ears!” approach as limiting the analyst to the discovery of a few superficial features and a few oppositions of contrast. He argues that auditory perception suffers from great individual variation, and there
are difficulties in fixing exact frames of reference for study. Absolute time, in seconds, may not relate to the musical conception of time in the piece. There are difficulties with timbral points of reference in attempting to describe sounds (pseudo-bell, pseudo-strings, etc) leading to oversimplification of compositional workmanship and loss of autonomy for new and unique sounds by being compared to a known quantity.

Stroppa argues that there is a choice, in instrumental music, between auditive analysis and analysis of the score. He implies there is no such choice in the analysis of tape music. Stroppa also asserts that: “Visual examination of the score (in instrumental music) encounters no time constraint, and may easily penetrate beneath appearances towards the generative ideas of the work.” (Stroppa, 1984: 180) In other words the score enables an analyst to examine details of a work, and such examination is not subject to real-time constraints.

How can we achieve some rigour in the systematic analysis of acousmatic music, and yet avoid the superficial findings that Stroppa warns us of? Let’s look at some previous attempts.

3.2 Some frameworks for the analysis of acousmatic music

Previous work has attempted frequency domain analysis of acousmatic music using representations known as MQ plots (Hirst, 1995). A McAuley and Quatieri (1986) analysis graph plots the tracks of significant frequency components (Y-axis) against time measured in analysis frames (X-axis). Amplitude information does not appear on MQ plots, but the threshold of significance can be varied.

The usefulness of MQ analysis lies in being able to isolate details within a work, as well as viewing the overall structure of a work. MQ analysis is a handy visual means for finding (and linking) similar musical events. So the use of MQ analysis can facilitate the isolation of foreground, middle ground, and background “streams” within a musical work. An MQ plot is able to function as a score in a descriptive sense, and thus it goes some way toward addressing Stroppa’s first criticism.
Stroppa's second problem concerns the difficulties of perception in determining frames of reference for time and timbre. The frame of reference for time used in previous analytical methods was absolute time in minutes, seconds, and hundredths of a second. This proved to be perfectly adequate for the limited exploration of temporal relationships undertaken.

The frame of reference for timbre remains an awkward problem. Resorting to metaphoric description still seemed more practical than Cogan's “sonic features”, or Smalley's “spectromorphology” (Smalley, 1986), which both seem too general. Furthermore they do not take account of the Gestalt effect of recognising a timbral unit within a complex spectrum, and the associations we make within the world of our own experience.

While previous research sought to find a useful description of the signal in order to try and catch a glimpse of some generative aspects of the work, it did not adequately address the socio-political, cultural, or psychological motivations, processes, or influences within a work. A higher level or higher order approach than signal level (or spectral plot) seems desirable.

The other extreme is to view the acousmatic sonic production as a work of art or a cultural object. The field of semiotics has created frameworks and methodologies that have proven quite useful in the analysis of some cultural artifacts, and are well summarized in the doctoral dissertation of Emmerson (1982).

Emmerson's approach to analysis can be subdivided into three stages. The definitive stage determines what is to be analysed. It sets the boundary conditions. The descriptive stage asks: “How does the system behave?” The final explanatory stage is concerned with why the system behaves as it does.

The first stage of analysis is the definitive stage and Emmerson's starting point is a model of the musical process characterised by Jean-Jacques Nattiez (1975) as material existence of the music “in itself” (the neutral level) sandwiched between two cognitive processes or “poles”. The poietic pole is that of production, the composer/creator's point of view. The aesthetic pole is that of reception, the listener's point of view.
According to Emmerson, analysis requires description and explanation, and these requirements necessitate the consideration of the idea of “score”. Emmerson defines an expanded view of “score” as not only the total determinants necessary to produce a piece of music, but also the total description necessary to explain a piece of music. For Emmerson, the neutral level of Nattiez (the material existence of the work) is both the score and its performance or realisation. However: “... the neutral level can only be described using a language involving aesthetic and cognitive components, and can only be explained in a hypothesis involving poietic considerations.” (Emmerson, 1982: 52)

Thus Emmerson’s revised model of the musical process moves from the poietic pole to the prescriptive score, to the acoustic signal, and finally the aesthetic pole - from production to reception.

Emmerson’s explanation stage of analysis provokes him to consider the notion of cause, and he groups four causes into two levels of dichotomy that seek to: “Account for the object as it exists.” and “Account for the reasons for the existence of the object.” (Emmerson, 1982: 81)

The examination of an acousmatic work as a cultural artefact seems to leave some gaps in connecting a work’s attributes with perceptual effect. A framework based on listener perception and cognition may provide a sound base upon which to build.

### 3.3 Developing a cognitive framework

#### 3.3.1 Why a cognitive model?

In adopting the perspective of the listener, we need to understand how the listener perceives a given work on a moment-to moment basis. There has been quite a lot of research into cognitive modelling of auditory information processing with regard to tonal music, but not acousmatic music. However, basing an acousmatic music framework on a related area where there has been some psychological research may prove as useful as deriving a generative grammar of tonal music based on linguistic research.
3.3.2 Bigand’s cognitive model for tonal music

Emmanuel Bigand provides us with a model of how one might go about developing a cognitive framework for the interpretation of tonal music. In an article curiously titled *Contributions of music to research on human auditory cognition*, Bigand (1993) classifies the body of work devoted to music perception into two main ways of conceiving music:

The first essentially stresses the highly complex acoustical and temporal nature of musical stimuli. For listeners, music is initially just an agglomeration of atmospheric vibrations that strike the eardrums. … The second approach deals with music as a complex sequential organization. The question here is how listeners manage to perceive relations between sound events separated in time. (Bigand 1993:231).

The article is an attempt to bring these two views together in a single approach. Bigand divides the approach into three parts: the characterization of the listener’s abstract musical knowledge structures; the way the knowledge is used and combined during the real-time processing of musical information (“event structure processing”); and finally “the symbolizing processes involved once the event structure has been analysed.” (Bigand 1993:233)

For Bigand, abstract musical knowledge takes two forms: “a system of relationships among musical categories (such as pitch categories, scale structures, and tonal and metrical hierarchies), and a lexicon of abstract patterns that are frequently encountered (such as the gallop rhythm, gap-fill melody, sonata or rag form)” (Bigand 1993:233) Listeners possess a knowledge of musical categories and structures, for the culture in which they belong, even though the knowledge may not be explicit. In Western tonal music, much of the attention is on the organization of pitch, and this has been the focus of much of the research in cognitive psychology.

3.3.3 Tonal Hierarchies

In trying to isolate what constitutes musical knowledge within listeners, Bigand focuses on tonal theories, and the discussion of pitch is further divided into melodic perception and harmonic perception. Listeners are able to discern changes in key and hierarchies within a single key.
Studies have shown that listeners are able to attribute a specific weight to each note within a tonal melody (Krumhansl, 1979; Krumhansl and Kessler, 1982; Bharucha, 1987) even if they are not musicians (Francès, 1988). Such studies have also led to the conclusion that these hierarchies are cognitive rather than sensory. Bigand notes that a consequence of hierarchy is the notion of interplay between tension and relaxation:

> These different hierarchical weights create differences in musical function – hierarchically important notes are notes on which the musical passage may end, whereas less important notes (from a hierarchical standpoint) are only passing tones. On the psychological level, these differences in function generate patterns of musical tension and relaxation. (Bigand 1993: 236)

Bigand points out that the existence of musical tension and relaxation, as a psychological phenomenon, has been demonstrated in the work of Krumhansl (1979) and Bharucha (1984).

An extension of this theory of melodic hierarchies is to create a further level of abstraction – a hierarchy of hierarchies:

> Since a hierarchy of structural importance exists, it might be thought that these patterns of extension and relaxation are also perceived in hierarchic fashion. This is one of the central hypotheses of the model proposed by Lerdahl and Jackendoff (1983). (Bigand, 1993: 237)

The discussion then turns to harmonic hierarchies, within a given “key”. Music theory tells us that chords (triads) built on each degree of the major scale are related in a hierarchical manner: “In various experiments, Bharucha and Krumhansl (1983) and Krumhansl and Kessler (1982) observed implicit knowledge of these hierarchies, even among listeners who had never studied music theory.” (Bigand 1993: 238)

Such harmonic hierarchies induce musical tension and relaxation, and further: “The psychological reality of such musical tension and relaxation as generated by harmonic hierarchies is confirmed by the phenomenon of asymmetry noted by Bharucha and Krumhansl (1983)” (Bigand 1993: 238)

Bigand reiterates the contention that levels of musical tension and relaxation are themselves organized in a hierarchical fashion.
What of harmonic relations between different keys? Bigand highlights a number of experiments that have demonstrated that listeners possess implicit knowledge of the distance between keys.

This knowledge of the distance between keys plays an important role in the mental representation of long musical fragments. Thus in a recent experiment I asked 40 musicians and 40 non-musicians to indicate the boundaries of the salient groupings within a long passage of chords (Bigand, 1991). (The) results show that listeners can temporary structure long musical passages on the basis of tonal hierarchies alone; musical training does not have a strong impact on this performance. (Bigand, 1993:240)

Listeners can travel within this multi-dimensional space created by inter-key and intra-key hierarchies, and this forms the basic framework of the tonal system.

3.3.4 Lexicon of schemas

The knowledge of pitch hierarchies is not the whole story however:

In every musical culture, rhythmic, melodic, and rhythmic-melodic configurations recur frequently enough for listeners to store them in long-term memory in the form of a lexicon of schemas and proto-typical forms. Some experimental research has revealed the existence of a lexicon in Western music culture. (Bigand, 1993: 242)

Some examples of these schemas can be found at the melodic level and the formal organizational level. A set of melodic patterns that underlie the basis of many tonal melodies have been identified by Meyer (1973) and Narmour (1983, 1989). Several notable examples are the “gap-fill melody” and the “changing note process”. At the formal level, research by Francès (1988) and Imberty (1979, 1981) “…has demonstrated, moreover, that Western listeners possess a knowledge (implicit or explicit) of various typical formal schemas comprising the temporal macrostructures of pieces of tonal music such as sonata form (exposition, development, recapitulation) and rondo form (couplets, refrain).” (Bigand, 1993:243)

These are essentially stereotyped macrostructures that are familiar to listeners. A structure such as AABA has empty slots that are to be filled during listening by details specific to the particular piece.
Bigand reminds us that there is evidence to show that the organization of temporal relations is culture-specific and that an African listener’s perception of a rhythmic passage may differ considerably from a Western listener’s perception. The scope of Bigand’s article is very much within the confines of Western tonal (art) music.

3.3.5 Event Structure Processing

Revealing the existence of an abstract musical knowledge, through schemas or hierarchies, does not explain how listeners perceive and analyse music, only how the information may be represented or stored. To discover what is happening in the real-time process of listening we must turn to the research into “event structure processing”.

Bigand starts with the theory proposed by Lerdahl and Jackendoff (1983) as the basis of a model for event processing. Their generative theory of tonal music was originally designed to describe the final stage of musical understanding not real-time event processing. However Bigand finds it a useful starting point.

Based on the generative theory of tonal music, Bigand begins with a three stage process as a foundation for the event structure processing model: time-span segmentation; time-span reduction; and prolongation reduction.

Lerdahl and Jackendoff’s theory is based on an ordered sequence of operations. “Prolongation reduction is possible only if the musical information has been organized in a hierarchic fashion. Yet this organization itself is only possible once the musical information has been segmented into groups.” (Bigand, 1993:246)

To the three stages derived from Lerdahl and Jackendoff, Bigand adds two preliminary stages provided by McAdams (1987): Reading the acoustic surface; and Auditory image formation.

The event structure processing model, as proposed by Bigand (1993), is summarized in Figure 3.1. Each stage will then be considered in detail.
Figure 3.1 Bigand’s model of event structure processing

**Reading the acoustic surface**

This is the process of transduction carried out by the peripheral auditory system to transform atmospheric vibrations into nerve impulses. The output of this stage is therefore nerve impulses.

**Auditory image formation**

This consists of two fundamental processes:

1. Simultaneous grouping of sound elements into a perceptual fusion that seems to emanate from the same source.

2. Sequential grouping of events originating from the same source (auditory stream formation and segregation).

There are some consistencies with Bregman’s approach here.

At the conclusion of the second stage, information is represented in the perceptual system as a set of notes, simultaneous or sequential, possessing qualities of pitch, timbre, loudness, and duration. This is the musical surface. “Truly musical
organization only begins to occur at subsequent levels of processing.” (McAdams, 1987:45)

**Time-span segmentation**

Bigand asserts that listeners segment a musical passage in time through two simultaneous processes: grouping according to Gestalt principles and metric structuring. We’ll discuss these in more detail, but first some introductory remarks.

When we listen to a piece of music, we don’t feel we are hearing a simple succession of notes. There is a certain ‘psychological present’ (or ‘perceptual present’) corresponding to a period of time during which auditory stimuli are held present for perception. It corresponds to the 4-5 seconds of short-term memory, and contains a limited number of elements (seven plus or minus two according to Miller (1956))

Chunking can organize elements into subgroups and therefore increase this number.

According to Bigand (1993:247):

> It means that the musical surface is apprehended through a series of ‘perceptual centrations’ within a window sliding along a string of sound events.

Segmenting the musical surface helps define the position and size of the sliding window. This is similar to the way we might read letters and words in a text.

Experiments have shown that these processes operate in a manner conforming to the grouping principles of Gestalt theories on form.

Deliège (1987), for example, conducted experiments to confirm the relevance of grouping principles based on proximity, similarity, continuity, and symmetry. A set of segmentation rules were constructed so that segmentation occurred according to the following criteria:

*Proximity rules:*

Rule 1: Slur or Rest

Rule 2: Attack point
Change rules:

Rule 3: Register

Rule 4: Dynamics

Rule 5: Articulation

Rule 6: Length

Rule 7: Timbre

In one experiment, subjects were asked to segment passages from the baroque, classical, and romantic repertoires. In a second experiment, a specially constructed nine-note stimulus that tested rule conflict was used. The main findings are summarized by Bigand (1993:250):

In general, the segments marked by subjects conformed to the ones predicted by the grouping rules. Those marked by musicians conform better than those by non-musicians. This difference being clear only in the first experiment. Listeners seem to prefer the proximity rules based on attack and timbre over other rules.

Alongside segmentation into groups, the detection of regular beats separated by a uniform time interval seems to be another process that operates within listeners. If some beats are detected as being more important that others then a metrical structure is perceived. A perceived regular ‘metric structure’ is demonstrated by a listener’s ability to ‘keep time’ with the music.

Metric organization is therefore an abstraction, an idealization, implying the existence of complex cognitive processes leading listeners to interpret a succession of sound events as being regular even though they are not strictly regular on an objective level (unless played by a computer). (Bigand, 1993: 252)

How far can metric hierarchy be extended? Is there a maximum time lapse beyond which the feeling of a periodic return of accent vanishes? Work by Fraisse (1974) suggests this period cannot exceed four to five seconds and thus metric organization would seem to be restricted to a local level within the organization of the musical surface.
The combination of the processes of grouping and metric abstraction leads to a segmentation of the musical surface into different time spans. Lerdahl and Jackendoff (1983) argue that that lower levels of segmentation are defined by metric structures while the higher levels are defined by grouping structures.

This third stage of processing leads to a representation in the form of groups of notes possessing an internal structure yet not exceeding the temporal limits of the psychological present.

It is obvious that if information processing stopped at this stage, music would be perceived as a juxtaposition of unrelated groups of notes. For that matter, this would seem to be the case with children up to the age of ten…Among adults, however, music is perceived in a different manner. It is therefore necessary to consider a fourth stage of information processing… (Bigand, 1993: 253)

**Time-span reduction (Establishing event hierarchies)**

In the next stage of the Bigand model, the need for the establishment of hierarchies is raised. The argument goes something like this: The need for a hierarchy stems from a need for information reduction which, in turn, stems from the limits imposed by the ‘psychological present’ (associated with short-term memory).

A corollary of this is that if long-term structures and relationships are to be created and perceived in acousmatic music, there must be some form of hierarchical organization operating. In regard to tonal music, let’s look at some of Bigand’s arguments in more detail.

Returning to the sliding window metaphor: “successive sound events are perceived through the frame of a window sliding along the sequence” (Bigand, 1993: 253-254). Information is lost when one segment ends and a new one begins. Any information passed over will only become subsequently available at the conceptual and symbolic levels. How can information in musical segments be coded into more abstract (and data-reduced) forms?

Experiments with words have shown that the ability to represent a long sequence of events with a hierarchy of operators is an economical way to code information from a sequence. In music, Deutsch (1980) argues that the high rates of recall of melodic
sequences show that listeners perceive and use hierarchies to memorize musical sequences. She showed that musical information can be represented in algorithmic fashion as a hierarchy of operators (in this case melodic operators). However the melodic segments were so short, the results could not necessarily be extended to all works in the repertoire. Bigand asks: “What types of operators are at work in real pieces of tonal music?” (Bigand, 1993: 255)

Bigand (1990) performed some experiments in which an initial group of musicians listened to a musical passage and were asked to transcribe the melody afterwards. A second group listened to a melody composed of the same note groups but played in a different order so as to render the tonal syntax harmonically incoherent. Test results showed that the number of correctly recalled notes was higher for the first melody. Bigand concluded:

> It would therefore appear that harmonic hierarchies established by the tonal system constitute a good way of summarizing musical information contained in relatively long melodies. (Bigand, 1993:256)

Experiments testing the recall of metric structure and harmonic structure (Sloboda and Parker, 1985) also suggest that the hierarchical representation of musical information contained in segments of a musical surface assists listeners in reconstructing relationships between those segments. This may seem an obvious point, but Bigand underlines its importance by quoting Meyer:

> Hierarchical structures are of signal (sic) importance because they enable the composer to invent and the listener to comprehend complex interreactive musical relationships. If stimuli...did not form brief but partially completed events (motives, phrases, etc.) and if these did not in turn combine with one another to form more extended, higher-order patterns, all relationships would be local and transient – in the note to note foreground. (Meyer, 1973: 80)

Bigand notes that the main working hypothesis of the cognitive sciences of music is the coding of musical information into an event hierarchy, but cautions that “this does not exclude the existence of other possible forms of representation – of an associative nature, for instance – that remain poorly understood today”. (Bigand, 1993: 257)

Information reduction can be summarized as organizing the musical surface into groups or segments and then comparing events within such groups for their ‘relative
stability’ (Lerdahl, 1989). The hierarchical relationships thus established provides an abstract representation of the musical structure.

Experiments were undertaken by Serafine et al. (1989) to investigate to what extent a listener is able to abstract the reduced structure underlying the musical surface. Musical fragments were played followed by four reduced structures representing foreground reductions and middleground reductions containing fewer notes each time. Two of the four reductions contained structurally insignificant notes. Subjects had to indicate which reduction fitted best. Their results showed that listeners were able to identify the correct structural reductions at better than chance levels. These results provide evidence to support the hypothesis that a hierarchical coding of musical information exists, but what are the exact cognitive processes involved in the coding operation?

Lerdahl and Jackendoff’s model emphasizes that the structural importance of an event depends on rhythmic and melodic features and the way they interact with the tonal hierarchy. A dominant falling on a strong beat, for example, will accentuate its structural significance. For Lerdahl and Jackendoff, rhythmic criteria supplement pitch criteria in the determination of the structural importance of events. To understand their model fully, tonal hierarchy must be distinguished from event hierarchy:

The former refers to an atemporal mental schema representing a system of culturally determined pitch relationships, whereas the latter refers to a structure that listeners must infer from the ongoing temporal sequence of musical events. (Bigand, 1993: 261)

In the Lerdahl model, ‘stability conditions’ represent tonal hierarchy and ‘time-span reduction’ represents the event hierarchy.

If this model of hierarchy creation is to be used for event structure processing, what processes activate knowledge of pitch hierarchy, and how can this information be combined with other sound parameters to define the event hierarchy?

To answer the first question, an early model of key identification (Longuet-Higgins, 1976) suggested that notes were represented in long-term memory in a three-dimensional psychological space where the key is defined as a subspace pictured as a
window. “When a listener hears a series of notes (C-E-G-A-B), the window is shifted to contain those pitches…The listener then decides, by taking other criteria into account, which key best corresponds to this passage.” (Bigand, 1993: 261) Other experiments have examined the impact on key recognition of other factors such as temporal order, duration and accent.

Once the key has been determined, how is the tonal weight of an event combined with other sound parameters to determine the event’s importance in the event hierarchy? Research by Palmer and Krumhansl (1987a,b) suggests that the rhythmic dimension contributes criteria to pitch hierarchies in determining the structural importance of events. Thus tonal grouping and metric structuring interact in the process of establishing event hierarchies. Bigand asserts that other sound dimensions such as loudness and timbre may play a similar role to rhythm.

In this fourth stage of processing (known as ‘time-span reduction’), event hierarchies are established by taking into account an event’s tonal weight and its rhythmic and metric value within a given group of notes. The tree diagrams “represent the hierarchical network of relations as well as the dominant musical event that encapsulates the information contained in those groups”. (Bigand, 1993: 263; also see Figure 8.8c p 248).

**Prolongation reduction (Perception of patterns of musical tension and relaxation)**

This fifth stage of processing provides a much more in-depth analysis and explores the relationships between events within groups, between groups within sections, and between sections within the entire work. Bigand notes that this form of processing concerns itself with goals, points of arrival, points of departure, and detours. “The common metaphor of ‘musical discourse’ clearly conveys this impression that the temporal organization of (tonal) musical (sic) is similar to the telling of a tale” (Bigand, 1993: 263). Another familiar metaphor is that of the narrative, which can include variations on the tension-problem, crisis, denouement pattern. Imberty and Bigand point out that “tonal syntax is partly responsible for this impression of the dynamic progression of musical flow. …tension is created when an element that is insignificant from the standpoint of syntax (and therefore of the tonal hierarchy) is inserted between two important events and vice versa.” (Bigand, 1993: 264)
Prolongation importance of an event also depends on other factors like the group in which it appears, its rhythmic value, and its metric position.

Lerdahl (1989) likens prolongation reduction to Schenkerian reduction in that it describes the linear continuity, departure, and return of events in hierarchical fashion, but it is more restricted than Schenker’s hierarchy. Prolongation connections derive from global to local levels of the associated time-span reduction. Lerdahl uses a top-down approach where designated events within a ‘prolongation region’ are evaluated in terms of their relative stability of connection by referring to a set of stability conditions. This approach is reiterated at each of the various levels. How this may operate within event structure processing is not clear. Lerdahl describes three main types of connection as follows:

Prolongation connections are represented by branchings in a tree diagram above the music, and … by slurs between noteheads in the musical notation. Right branches stand for tensions motion (or departure), left branches for relaxing motion (or return). The three kinds of connection are strong prolongation, in which an event repeats; weak prolongation, in which an event repeats in an altered form (such as triadic inversion); and progression, in which an event connects to a completely different event. (Lerdahl, 1989: 71; see also Figure 4, p 72)

The final output of this fifth stage produces a hierarchical network of tension and relaxation.

Symbolization processes

Examines how the output of event structure processing can be interpreted at a symbolic level. This stage is probably the most complex, and yet least researched. Discussion of this stage is so extensive that we must set it aside to be the subject of its own chapter.
3.4 A cognitive framework for acousmatic music

3.4.1 Basic Assumptions

The idea of adapting a cognitive model for the processing of Western tonal music to acousmatic music implies some inherent assumptions. Firstly it assumes that a hierarchical means of organization and mental representation operates.

One can fairly readily argue that time span segmentation may also operate in acousmatic music processing and its output would be groups of sound events with internal structure. If information processing stopped at this stage then acousmatic music would be perceived as a juxtaposition of unrelated groups of sound events. Are all relationships local and transient in the event to event foreground of acousmatic music?

If long-term structures and relationships are to be established and perceived in acousmatic music, then there must be some form of hierarchical organization operating. This need for a hierarchy stems from a need for information reduction resulting from the limits of the psychological present (short-term memory). Further “evidence” of hierarchical organization would be whether there is a perceptible scheme of alternating periods of tension and relaxation, or discourse, as has been established in tonal music. On the other hand, organisation could be associative or borrowed from some extra-musical experience, and could be quite “flat” structurally. If there are some kind of “event hierarchies” in acousmatic music, then what features within the music establish such hierarchies?

Another assumption in this discussion will be that metric organization is not as significant for the type of music we want to consider. In other words we will select works in which a basic pulse is not discernable or a major factor in the unfolding of the work. This is certainly a characteristic of many acousmatic pieces. We are seeking to hold one variable constant in order to consider whether hierarchical structuring is possible when sound events (or timbral units) are the primary structuring force. Structuring in time will still be important, but we will initially consider works where no apparent pulse is present.
3.4.2 Event Structure Processing in Acousmatic Music

Bigand’s article divides the cognitive approach into three parts: the characterization of the listener’s abstract musical knowledge structures; the way the knowledge is used and combined during the real-time processing of musical information (event structure processing); and finally the symbolizing processes involved once the event structure has been analysed. Only the first two parts will be addressed in this chapter.

Firstly there needs to be a consideration of what constitutes an “event” in acousmatic music, and what constitutes a listener’s musical knowledge within the acousmatic “culture”. Further, what syntactic system is capable of producing some form of organization in the absence of the tonal system?

Bigand presents a model of event structure processing whereby a ‘sliding window’ of listener perception analyses the work on a moment by moment basis, and a matching process compares salient features with musical knowledge representations in long-term memory. It implies that where relevant knowledge representations are recognized, or activated, this activation is somehow retained for processing at some time in the future.

For a listener processing ‘musical’ auditory information, Bigand’s model of event structure processing identifies six stages: reading the acoustic surface; auditory image formation; time-span segregation; time-span reduction; prolongation reduction; and symbolization processes. Let’s examine how appropriate each of these stages might be for acousmatic music.

**Reading the acoustic surface**

This is the process of transduction carried out by the peripheral auditory system to transform atmospheric vibrations into nerve impulses. The output of this stage is nerve impulses.

This stage would remain unchanged for acousmatic music.
Auditory image formation

This consists of simultaneous grouping of sound elements into a perceptual fusion of events emanating from the same source, and sequential grouping of events originating from the same source. At the conclusion of this second stage, information is represented in the perceptual system as a set of notes, simultaneous or sequential, possessing qualities of pitch, timbre, loudness, and duration. So for tonal music, this set of ‘note events’ represents the ‘musical surface’.

In the case of acousmatic music, we must examine this second stage much more closely since it is not so clear cut what corresponds to the tonal notion of ‘note event’. In acousmatic music, environmental sounds may be included, along with processed environmental sounds, recorded instrumental sounds, processed instrumental sounds, synthesized sounds, and hybrids of these types.

The sheer number of possible combinations of such a wide variety of sound sources is far greater than the number of traditional combinations within Western tonal music. Arguably, the acousmatic music listener is confronted with a large number of novel works with unusual combinations of sounds, and consequently the task of recognizing sound events becomes an important consideration.

McAdams (1993) has written a detailed account of sound source/event recognition. The stages he characterizes are: Sensory Transduction; Auditory Grouping (simultaneous &/or sequential grouping of unrecognized sound events); Analysis of Auditory Features (abstract properties characterizing sound source invariants); Matching with Auditory Lexicons (lexicons of meanings, significance, and names); and finally Recognition (See Figure 3.2). He also emphasizes the feedback loops expressed as schema-driven (top-down) processes that enable higher order cognitive functions to influence lower order functions, eg. recognition of some elements in an auditory stream may assist feature extraction of other elements which contributes to their recognition.
Figure 3.2 McAdam's Stages of Processing in Auditory Recognition

The output at the conclusion of this second stage for acousmatic music may be a set of ‘sound events’, simultaneous or sequential, possessing qualities of timbre, loudness, duration, degree of pitch (ranging from definite pitch centroid to undefined), and recognized sound source (real, surrogate, fictional).

An important additional element for acousmatic music, that may be associated with auditory image formation or time-span segregation or as a distinct entity of its own, is what I shall define as ‘auditory scene recognition’
Auditory scene recognition

The phrase ‘auditory scene recognition’ has a certain resonance with Bregman’s notion of ‘auditory scene analysis’. The term ‘scene’ is also congruent with the genres of film and drama in that it has connotations of a certain place and time.

Bigand comments that when we listen to a piece of music, we don’t feel we are listening to a simple succession of notes, but rather there is a certain psychological or perceptual present corresponding to a period of time during which auditory stimuli are held present for perception. This provides a hint that we should look at the associations and relationships between recognized sounds within the psychological present.

Auditory scene recognition may involve the perception of such relationships so that the listener comes to recognize a certain sound environment. It is perception of the sound environment that helps to place us within the world, in addition to any visual or other stimuli, and this may be a strong perceptual inclination. The establishment of changing auditory scenes is an important element in the composition of acousmatic music, especially in totally immersive sound environments like headphone listening or multi-channel speaker diffusion.

When the auditory stimulus is acousmatic music, auditory scenes may seem ‘real’ or ‘fictional’ according to the disposition of sound events within the ‘sliding window’ of our perceptual present. Such categories may form the beginnings of timbral syntax that will need to be developed further.

In McAdams’ work on sound source and sound event recognition, he makes the distinction between recognition and identification (McAdams, 1993: 147-148). Recognition means correspondence with something that has been heard in the past, and may involve a sense of familiarity. Identification is a more narrowly focused form of recognition involving naming, or labelling, by accessing a lexicon of names. We can make the same distinction with auditory scene recognition and identification so that we may not be able to name a scene, but we may recognize it or recognize that it is like something within past experience. A completely ‘fictional’ scene may cause sensory confusion or may result in a new category being stored in memory.
The output at the conclusion of auditory image formation and auditory scene recognition, according to our postulated model for acousmatic music, would be a set of ‘sound events’ possessing certain qualities, and a recognized auditory scene which may be ‘real’, ‘fictional’ or some combination of the two. The categorization of a sequence of auditory scenes would be like keeping a running record of our presence within the sound world, as that world is changing. Abrupt changes in the sound world, like opening a door upon a street, signal implications for time-span segregation of the work.

McAdams (1993) highlights the need for further research in the recognition of non-instrumental, non-verbal sounds and points out that much of the research so far has concentrated on the recognition of isolated sounds. Our discussion points to the need for more research on auditory scene recognition, which may explore the association and interaction of many simultaneous and sequential sound events.

**Time-span segregation**

Bigand asserts that listeners segment a musical passage in time through two simultaneous processes: grouping according to Gestalt principles and metric structuring. Since we are putting metric structuring aside for the time being, we will only consider grouping according to Gestalt principles for acousmatic music here.

As we saw for tonal music, segmenting the musical surface helps define the position and size of the sliding window of perception. This is similar to the way we might read letters and words in a text. Tonal music experiments by Delière (1987) have shown that these processes operate in a manner conforming to the grouping principles of Gestalt theories based on proximity, similarity, continuity, and symmetry. As we saw earlier, a set of segmentation rules were constructed so that segmentation occurred according to the following criteria:

**Proximity rules:**

- Rule 1: Slur or Rest
- Rule 2: Attack point
Change rules:

Rule 3: Register
Rule 4: Dynamics
Rule 5: Articulation
Rule 6: Length
Rule 7: Timbre

How should the above grouping principles by adapted and augmented for acousmatic music?

Proximity rules

Rule 1 is relevant and could be adapted from ‘Slur or Rest’ to ‘Stream or Silence’ so that sequential sound events would tend to be grouped together representing an auditory stream whereas a silent passage would define a boundary between groups.

Rule 2 could be modified from ‘Attack Point’ to make use of the computer music term ‘Inter-onset Time’. Where the inter-onset time between sound events is zero (for simultaneous events) or comparatively small, those events will tend to be clustered together in the same group. An extreme example of this phenomenon is in the case of so-called granular synthesis where the grouping of separate events becomes fused into a single event.

Change rules

Rule 3 also applies, but in this case abrupt changes in register would be applied to frequency rather than pitch. To reinforce the distinction let us rename this rule ‘Frequency Register’.

Rule 4 ‘Dynamics’ is also a segmentation rule for acousmatic music, but use of the term ‘Amplitude Profile’ will allow us to identify segments according to significant contrast in the overall dynamic of several sound events and significant differences in
the amplitude envelope of two sound events (eg. A loud sound with sudden decay will contrast with a loud sound with a lengthy sustain over the same duration).

Rule 5 ‘Articulation’ should be renamed to ‘Excitation Method’ to reflect the broader scope of sounds present in acousmatic music.

Rule 6 ‘Length’ also applies to acousmatic music but is more appropriately termed ‘Duration’ since sound events with substantially different durations will tend to be grouped apart.

Rule 7 ‘Timbre’ is very relevant but may require a more sophisticated approach. Breaking this down further we find that timbre is related to the method of excitation and the resonant properties of the body excited (Vanderveer, 1979). Here there is some overlap with Rule 5’s ‘articulation’ and Rule 4’s ‘dynamics’, and perhaps even the rest of the rules as well! Since timbre is spectrum-related, then we could coin the phrase ‘Spectral Profile’ to identify grouping where sound events have similar spectral profiles and segregation where sound events display contrasting spectral profiles (eg. Harmonic-inharmonic).

Additional rules for acousmatic music include:

Rule 8 is an additional segmentation criterion which overlaps with Rule 7 ‘Spectral Profile’ but is significant enough in the genre of acousmatic music to attract its own listing. We will call it ‘Source-cause’ (after Smalley, 1994) to reflect the fact that sound events which activate similar real-world sound sources or causes will be grouped together (eg. Chicken sounds versus cow sounds). Sound event source-causes may be merely recognized (un-named) or identified (named). Where no real-world sources or causes are recognized then the cognitive processes may seek a ‘surrogate’ in the case where a similar category is activated in long-term memory, or a new category may be established.

Rule 9 is an addition that stems from the discussion regarding auditory scene recognition above. Where a particular collection of sound events changes to a substantially different collection of sound events, there is segmentation between the two contrasting ‘auditory scenes’.
Our modified list of segmentation rules for acousmatic music has become:

*Proximity rules:*

- Rule 1: Stream or Silence
- Rule 2: Inter-onset Time

*Change rules:*

- Rule 3: Frequency Register
- Rule 5: Excitation Method
- Rule 6: Duration
- Rule 7: Spectral Profile
- Rule 8: Source-cause
- Rule 9: Auditory Scene

The process of grouping leads to a segmentation of the musical surface into different time spans.

This time-span segregation creates a representation in the form of groups of sound events possessing an internal structure yet not exceeding the temporal limits of the psychological present.

*Time-span reduction*

As we saw above with regard to tonal music, in the next stage of processing known as ‘time-span reduction’, event hierarchies are established by taking into account an event’s tonal weight and its rhythmic and metric value within a given group of notes. Tree diagrams are used to represent the hierarchical network of relations as well as the dominant musical event that encapsulates the information contained in those
groups” (Bigand, 1993: 263; also see Figure 8.8c p 248). Is there any equivalent to ‘tonal weight’ in acousmatic music?

In an article on timbral hierarchies, Lerdahl cautions that:

…metrical structure and time-span reduction – do not seem to suggest anything toward building a timbral hierarchy (Lerdahl, 1987:138)

However Lerdahl is considering pure, abstract musical events only. Acousmatic music’s inclusion of concrete sound provides a basis for hierarchical perception on a mimetic-aural continuum. Lerdahl does go on to develop a timbral hierarchy using grouping structure and prolongation reduction.

Let us return to the sliding window metaphor where successive sound events are perceived through the frame of a window sliding along the sequence. Information is lost when one segment ends and a new one begins. Any information passed over will only become subsequently available at the conceptual and symbolic levels. How can information in musical segments be coded into more abstract (and data-reduced) forms in the case of acousmatic music?

Experiments with tonal music showed that the high rates of recall of melodic sequences indicated that listeners perceive and use hierarchies to memorize musical sequences. Another set of musical transcription experiments showed that the number of correctly recalled notes was higher for the melody which was ‘harmonically’ correct. Recall Bigand’s conclusion that “…harmonic hierarchies established by the tonal system constitute a good way of summarizing musical information contained in relatively long melodies.” (Bigand, 1993:256) Are there similar recall experiments that could be devised for acousmatic music?

Time-span reduction for tonal music involves information reduction. Information reduction can be summarized as organizing the musical surface into groups or segments and then comparing events within such groups for their relative stability. The hierarchical relationships thus established provides an abstract representation of the musical structure.
Experiments were undertaken by Serafine et al. (1989) to investigate to what extent a listener is able to abstract the reduced structure underlying the musical surface. Musical fragments were played followed by four reduced structures representing foreground reductions and middleground reductions containing fewer notes each time. Subjects had to indicate which reduction fit best. Are there similar ‘reduction’ experiments that could be devised for acousmatic music in order to establish whether there is a hierarchical coding?

Lerdahl and Jackendoff’s model emphasizes that the structural importance of an event depends on rhythmic and melodic features and the way they interact with the tonal hierarchy. Above we noted that for Lerdahl and Jackendoff, tonal hierarchy refers to an atemporal mental schema representing a system of culturally determined pitch relationships, whereas the event hierarchy refers to a structure that listeners must infer from the ongoing temporal sequence of musical events. In the Lerdahl model, ‘stability conditions’ represent tonal hierarchy and ‘time-span reduction’ represents the event hierarchy. In essence there is a matching process going on between the tonal hierarchy coding category contained in the listener’s lexicon of representations and the events as they unfold within the defined segments of the work.

For tonal music it was suggested that tonal hierarchy is established through a two-stage process of key identification followed by the attribution of tonal weight. An early model of key identification (Longuet-Higgins, 1976) suggested that notes were represented in long-term memory in a three-dimensional psychological space where the key is defined as a subspace pictured as a window. Once the key has been determined, the rhythmic dimension contributes criteria to pitch hierarchies in determining the structural importance of events (Palmer and Krumhansl, 1987a,b).

If we are to use this model of hierarchy creation, are there ‘stability conditions’ for acousmatic music, and what processes activate knowledge of sound event hierarchy?

Restating these questions in a more general way we can ask: Is it possible to identify some sound events in acousmatic music that are more structurally important than others? If so, then by what means is it possible to make such identifications?
Firstly we can state that a diminution of the reliance on pitch for structural organization will mean that other musical parameters will become more important. A sound event will be more prominent if it has one or more features which are extreme compared to its neighbours, eg. Long duration, loud dynamic, extreme frequency, sudden attack. In addition to single events, sounds may be interpreted in a collection as a ‘sound event complex’ and can be treated as a single entity. Without doubt though, the most defining characteristic in acousmatic music is timbre, so do timbral hierarchies exist in our perceptual system?

More experimental research needs to be done in this regard, but to inform such experimental design, what could one look for?

It seems that a dominant feature of our perceptual system is the desire to attribute a known source or cause for a sound event. In the previous chapter I attempted to amplify Smalley’s notions on defining and refining timbre (Smalley, 1994). Smalley defines the term ‘source bonding’ to encapsulate “the natural tendency to relate sounds to supposed sources and causes, and to relate sounds to each other because they appear to have shared or associated origins” (Smalley, 1994:37). Smalley explains that source bonding is extrinsic in that it refers to experiences outside the work itself. He then defines the term ‘extrinsic matrix’ to describe links to a wide range of real and imagined sounding and non-sounding phenomena that exist outside the work. Smalley creates a pseudo-timbral hierarchy relating to instrumental source-causes by describing a series of ‘source-cause levels’ divided into imminent (eg. the violin in this work), cumulative (violins in other works), extended (viol family), and dispersed (all string instruments in all cultures). He notes that in acousmatic music we do not find such a definable hierarchical basis for establishing the source-cause aspect of timbral identity, but he does seek to refine the intrinsic and extrinsic forces that help to shape timbral identity through ‘coherence’ and ‘strength’. The latter concepts will be explored more in the section on prolongation reduction.

Emmerson (1986) has also created a coarse-scale timbral hierarchy through his depiction of a language grid as it relates to the language of electroacoustic music. The language grid was really developed in order to classify whole works by allocating them to an individual cell of a 3x3 matrix. One axis of this matrix depicted the type of
musical syntax being used in the work (abstract or abstracted), and the other axis depicted a type of musical discourse according to the degree of mimetic reference used (aural discourse dominant or mimetic discourse dominant). It is this latter scale that may prove useful here.

We are now in a position to attempt to define a timbral hierarchy in terms of the degree of perceived source-cause a single sound event possesses and the degree to which it relates to other sound events and the auditory scene as a whole. We could devise a table extending from the least significant timbral weighting (1) to the most significant (6) as shown in Table 3-1.

<table>
<thead>
<tr>
<th>Weighting</th>
<th>Source-cause recognized?</th>
<th>Relationship with other events?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>2</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>3</td>
<td>Surrogate</td>
<td>No</td>
</tr>
<tr>
<td>4</td>
<td>Surrogate</td>
<td>Yes</td>
</tr>
<tr>
<td>5</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>6</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

**Table 3-1: Timbral Hierarchy Weightings**

In this table we have introduced the notion of a ‘surrogate’ source-cause, borrowed from Smalley (1994). The use of the term surrogate means that a perceived sound event may not have a recognized source-cause, but it may be similar to a recognized source-cause (eg. a bird-like sound). We could expand this table further by recognizing that when a sound event has a close relationship with another event, we are really speaking about spectral similarity between the events. At one end of the scale we would have repetition, at the other extreme there is complete dissimilarity, and in between there is transformation. Whether this scale should plot the nature of the relationship or the number of relationships a sound event has is open to further speculation and analysis.
By considering the relationships between sound events here, we are really overlapping with the next section on prolongation. It would appear that it is harder to erect boundaries in acousmatic music than it is in tonal music, and we shall expand on this further in the next section.

**Prolongation Reduction**

As we saw above, the next stage of processing involves the coding of ‘musical discourse’ or ‘narrative’. Such a discourse is sometimes characterized by the tension-problem, crisis, denouement pattern.

Lerdahl (1989) defines the three kinds of prolongation connection between events as: *strong prolongation*, in which an event repeats; *weak prolongation*, in which an event repeats in an altered form; and *progression*, in which an event connects to a completely different event.

It was pointed out that, for tonal music, tonal syntax is primarily responsible for the impression of the dynamic progression of musical flow of tension and relaxation, but prolongation importance of an event also depends on other factors like the group in which it appears, its rhythmic value, and its metric position.

In the realm of acousmatic music, Smalley writes that we should first of all establish identity, and then discriminate the ‘incidental’ from the ‘functional’ when examining spectromorphological evolutions.

Establishing identity in acousmatic music is not as clear cut as is the case with instrumental music. In acousmatic music, a timbre may be a discrete object, easily separated from its context or it may be a continuity intertwined with other continuities and not so easily separated. Smalley introduces the concept of ‘timbral level’, which concerns the relationship between the two continuous entities - duration and separability. An ‘event’ may be a short-term entity or a long-term evolution, and its separability can be defined as to whether it is a discrete object or a continuing context:

... Timbral level in traditional note-based music is quite simple. The note is the lowest level and is articulated by an instrumental source. Form develops from note articulations. In electroacoustic music continuing contexts resist and deny low-level segmentation. Thus once timbral level ceases to be clearcut we cannot separate
timbre and discourse: timbral attributes become woven into the spectromorphological fabric. (Smalley, 1994:43)

To help add some clarity to the discussion, Smalley introduces the concept of ‘coherence’:

The notions of existence and registration imply a certain coherence, otherwise identity would not be feasible. (Smalley, 1994:41)

Coherence in the case of instruments is usually associated with spectral fusion, which in turn is associated with harmonicity. Smalley prefers the term ‘integration’ to ‘harmonicity’, and this allows him to speak of an integration-disintegration continuum:

Integration means that within a sonic physiognomy the distribution of spectral components in spectral space, and their behaviour over time should not be such that a component or sub-group of components can be perceived as an independent entity. (Smalley, 1994:41-42)

While Smalley has highlighted the problem of discriminating between integration and disintegration, his solution to creating connections between identities in electroacoustic music is to introduce the concept of discourse:

Playing with integration and disintegration is at the very heart of electroacoustic musical discourse, a discourse which becomes spectromorphology itself once the timbre complex is spilt open, a discourse where the notion of timbre can at one moment perhaps be grasped, but at the next it evaporates. (Smalley, 1994:42)

Smalley specifies six types of discourse in acousmatic music. The first three types could be categorized as primary, and they are: source-cause discourse, transformational discourse, and typological discourse. The remainder can be described as secondary and concern the combinations of and relationships between the primary discourses: motional, behavioural, and tensile.

The primary discourse types will be explored, in the first instance. Note that the boundaries between these types overlap. Is there a hierarchy between the primary discourse types? Intuition seems to indicate that source-cause discourse will be dominant, and that transformational and typological discourses will be subordinate. All three discourses will interact with each other to varying degrees.
Source-cause discourse is concerned with the bonding play of specific or inferred sounding identities. Source-cause discourse will also include consideration of the connection of events within the auditory scene.

Smalley’s second type of discourse is typological discourse.

Identities are recognised as sharing timbral qualities but are not regarded as being descendants of the same imminent identity – they do not possess a common identity base. Typological discourse is associative. (Smalley, 1994:44)

Smalley invents the term ‘generic timbres’ to describe larger groupings of timbres. He gives two examples of generic timbres: noise and inharmonicity.

Transformational discourse concerns the situation where an identity is transformed while retaining significant vestiges of its roots. To achieve this, certain attributes must remain stable while others vary.

As noted in the previous chapter, Smalley makes some important points in passing regarding memorability and contiguity.

It is undoubtedly true that pitch relations and source-causes are the most easily memorable sonic phenomena, and that the problem with multiple spectrological attributes is that we do not know which ones are to be relevant in the discourse. This is why electroacoustic music which avoids pitch phenomena and strong source-cause references will be most frequently concerned with contiguous relationships. … Non-contiguous discourse can only be highly developed within the precise and detailed memorability of culturally imbedded pitch and rhythm systems. (Smalley, 1994:44)

Although Smalley elevates transformation to its own discourse category, one could argue that transformation is an ‘operation’ that is used within the other categories, and is not a separate category of its own. We hear transformation operating on source-causes and on spectral types. Transformation is the embodiment of Lerdahl’s weak prolongation connection type.

Acousmatic music discourse can therefore be distilled into two main emphases: source-cause discourse and typological discourse. Transformation is recognized as a fundamental operation for navigation through the multi-dimensional source-cause and typological timbre spaces. Source-cause timbre space will involve a high degree of
semantic meaning. Other musical discourses such as pitch discourse, rhythm/meter discourse, and the interplay of dynamics and other musical parameters certainly influence tension and relaxation, but they are subordinate to the source-cause and typological discourses. These results are summarized in Table 3-2.

<table>
<thead>
<tr>
<th></th>
<th>Tonal Music</th>
<th>Acousmatic Music</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dominant discourses</strong></td>
<td>Tonal</td>
<td>Source-cause</td>
</tr>
<tr>
<td></td>
<td>Rhythmic/metric</td>
<td>Typological</td>
</tr>
<tr>
<td><strong>Subordinate discourses</strong></td>
<td>Timbral</td>
<td>Pitch centric</td>
</tr>
<tr>
<td></td>
<td>Dynamic</td>
<td>Rhythmic/metric</td>
</tr>
</tbody>
</table>

**Table 3-2: Comparison of tonal and acousmatic discourses**

Having identified the issues relevant to prolongation reduction for acousmatic music in this chapter, further work is required to characterize tension and relaxation in source-cause discourse and typological discourse. Several attempts at the specification of a typological discourse, such as those by Lerdahl (1987) and Slawson (1985), may provide some insights. Source-cause discourse is an equally difficult area and will necessarily involve an examination of symbolization processes.

Ultimately the final output of this stage of event structure processing produces a hierarchical network of tension and relaxation (See Figure 3.3). The characterization of source-cause discourse also draws symbolization processes into the discussion as an overlapping or simultaneous stage of processing, but consideration of such symbolization processes requires its own special discussion.

Whether hierarchical structures are operative within acousmatic music also needs to be tested, and if they are present, how important are they compared with alternatives such as associative organisations?
3.5 Conclusion

In an attempt to examine the process of listening to acousmatic music as a cognitive experience, a model for the ‘event structure processing’ of tonal music has been used as the foundation of a framework. This chapter is purely speculative in that some modifications and additions to the Bigand model have been suggested, but without any experimental basis. Along the way, areas for psychological experimentation have been identified, and the important areas for consideration with regard to acousmatic music have been highlighted.

The next chapters will attempt to define tension-relaxation schemas for source-cause discourse and typological discourse, and will examine the relationship with symbolization processes. The concept of auditory scene creation and recognition is emerging as an important issue, with ‘real’ and ‘fictional’ sound events establishing relationships within ‘real’ and ‘fictional’ auditory scenes.
An important additional consideration will be the development of a suitable notational scheme as a tool for analysis and as a means of documenting and communicating the analytical findings.
Chapter 4 : Signs and Symbols, Syntax and Semantics

4.1 Introduction

Not only do we recognize sounds, but we ascribe meanings to sounds and meanings to
the relationships between sounds and other cognate phenomena. Music is a
meaningful and an emotional experience. In acousmatic music there are perceived
relationships between the sonic attributes of sounds. What is this “syntax” of
acousmatic music and how does it interact with the semantic references afforded by
some of the sonic material in acousmatic music?

To investigate these questions we will examine the relevant literature that details some
of the processes that may operate in the recognition of environmental sounds and also
consider processes involved in the perception of Western tonal music. Through
combining these two bodies of previous work, where there has been some research
activity, we may shed some light on the perception of acousmatic music, where there
has been only a meagre amount of research.

4.2 Signs, symbols and emotions in tonal music

The chapter on Emotion and Meaning from the text by Dowling and Harwood (1986)
provides a summary on the interpretation of signs and symbols with regard to Western
tonal music. An important distinction they make is between the representation of
emotions and the elicitation of emotions. Music can “represent emotional processes as
well as induce them … often the emotion represented is also the emotion induced,
though this is not always the case.” (Dowling and Harwood 1986: 203)

As their starting point, Dowling and Harwood use the classification system for signs
developed by nineteenth century philosopher Charles Pierce (1931-1935, Vol. 2).
Signs represent things or events, other than themselves, and Pierce describes three
types of sign: index, icon, and symbol. Although musical events are complex and
involve more than one sign function at a time, there is a special relationship between
musical events and symbolic representation: “We suggest that many of the meanings
of musical events depend on this symbolic mode of representation and that musical
symbols often depend for their meaning on being embedded along with other musical symbols in a structure of musical syntax.” (Dowling and Harwood 1986: 203)

Each type of sign will now be considered in detail.

4.2.1 Index

According to Dowling and Harwood: “Indexical representation involves the direct association of a musical event with some extra-musical object or event, so that emotions previously associated with the extra-musical object come to be associated with the music.” (Dowling and Harwood, 1986: 204)

The authors give the examples that Tchaikovsky’s use of cannon shots in *Overture 1812* or the fire sirens in Varèse’s *Poème Electronique* can evoke previous emotional reactions to cannons or sirens within the context of a musical work, although there is nothing particularly musical about the process. It is simply a conditioned response based on a previous, extra-musical, association.

4.2.2 Icon

Dowling and Harwood define iconic representation in the following way: “With iconic representation we turn to effects that depend on patterns within the music itself. Music can represent emotions iconically because the ebb and flow of tensions and relaxations in the music mirror the form of emotional tensions and relaxations.” (Dowling and Harwood, 1986:205)

The authors point out that individual musical forms can seem to bear both sad and happy interpretations, and that iconic signs require some context for their disambiguation. While listeners find it quite natural to attach emotional labels to pieces of music (Francès, 1988, Experiment XIV), Dowling and Harwood caution that listeners’ verbal responses cannot be taken as referring literally and precisely to emotional states. They quote Imberty to emphasize their point: “It is exactly the meaning that when explicated in words gets lost among the verbal significations – too precise and too literal – and gets betrayed … Music doesn’t signify, it suggests; that is it creates forces in the imagination that stimulate and orient verbal associations” (Imberty, 1975: 91)
While we should keep this caveat in mind, there is some evidence that listeners are consistent in the emotional characterization of some musical excerpts. Table 4-1 provides a summary of the work of Hevner (1935, 1936) who attempted to link the emotional properties of pieces with particular iconic features within the music by presenting different versions of the same piano piece and asking listeners to select suitable adjectives from a list.

<table>
<thead>
<tr>
<th>Musical Feature</th>
<th>Emotional Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Richness of harmony (Simple vs Complex)</td>
<td>Simple – merry, joyous, gay, playful</td>
</tr>
<tr>
<td>Mode (Minor vs Major)</td>
<td>Minor – pathetic, doleful, sad and dreamy, tender, yearning</td>
</tr>
<tr>
<td>Rhythmic character (Firm vs Flowing)</td>
<td>Firm – spiritual, lofty, dignified and vigorous</td>
</tr>
<tr>
<td>Melodic direction (Ascending vs descending)</td>
<td>Had the least effect</td>
</tr>
</tbody>
</table>

Table 4-1: Hevner's emotional interpretations of musical features

Wedin (1972) used multi-dimensional scaling techniques to analyze listeners’ ratings of 40 musical excerpts using a checklist of 125 adjectives. The responses seemed to cluster into three dimensions. The independently characterized adjectives and musical features are show in Table 4-2.
Table 4-2: Wedin's analysis of listeners' adjectives ascribed to musical features

These categorizations are fairly broad-brush. Are finer emotional discriminations possible? Brown (1981) conducted studies that didn’t rely totally on verbal responses but rather listeners were asked to match pieces with similar moods and then match with the appropriate adjective cluster. Brown also conducted a second study that was more fine-grained and he called it *Twelve Variations on Sadness*. Different groups of listeners were presented the task either with or without his pre-determined adjective pairs. Without words, both musicians and non-musicians achieved only moderate consistency, and they didn’t respond to the emotional nuances present. Dowling and Harwood contend that this highlights the importance of context in disambiguating emotional meaning of iconic representations:

The importance of context in aiding listeners to decipher iconic emotional meanings in music is shown in the with-words condition, where Brown also gave listeners the six adjective clusters in terms of which to categorize the pieces. In that condition, the non-musicians behaved just as in the without-words condition. … However, the musicians group made modal choices that exactly matched the original pairings. … Thus with added context of the suggested adjective clusters, musicians, but not non-musicians, were able to solve the task of retrieving the

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Adjective Characterization</th>
<th>Musical Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Contrasts energy vs relaxation</td>
<td>Energy – staccato (vs legato), loud (vs soft)</td>
</tr>
<tr>
<td>2</td>
<td>Contrasts gaiety vs gloom</td>
<td>Gaiety – consonant harmonies, flowing (vs firm) rhythm, major mode, relatively high pitch</td>
</tr>
<tr>
<td>3</td>
<td>Trivial and light vs solemn and serious</td>
<td>Serious (associated with cultural labels serious or old) – loud, slow, low in pitch, and avoiding major mode.</td>
</tr>
</tbody>
</table>
underlying meanings at better than chance accuracy. (Dowling and Harwood, 1986: 212)

Context can be readily provided by a story in opera or ballet music whereas, in other non-program music, context may need to be gleaned through say familiarity. Such familiarity may be achieved through repeated listening during rehearsal (musicians) or through repeated exposure via recordings (non-musicians).

4.2.3 Symbol

Dowling and Harwood define a symbol as:

… a sign that derives its meaning from its relationships in a network with other signs. Where musical signs are indexical or iconic, the meaning can often be read from the sign directly… The meaning of musical symbols arises from their place in the syntax of the piece and, more broadly speaking, of a style. This is true of some iconic signs as well. The play of tension and release over time that mirrors the ebb and flow of emotional excitement depends on syntactic relations for its expression, and so such icons also functions as symbols. … A symbol thus depends for its significance on its place in the musical pattern in relation to other symbols. (Dowling and Harwood, 1986:213)

Some critical questions now arise: How do patterns that link symbols together in music and cognition function in relation to emotion? How does the interpretation of a pattern lead to the generation of emotional meaning? How is emotion induced directly rather than its representation?

To help answer these questions, Dowling and Harwood’s cognitive theories of emotion will be recounted.

4.2.4 Emotion : A cognitive theory

Dowling and Harwood quote Mandler’s theory of emotion as a basis for their own discussion:

Human cognition operates by means of perceptual-motor schemata through which (largely unconscious) expectancies are generated for upcoming events and by which future behaviors are planned. The interruption of an ongoing schema or plan brings about biological arousal – a signal that something has gone wrong. This reaction in turn triggers a search for a cognitive interpretation of what happened – a search for meaning. The arousal and the interpretation join together in producing an emotional experience of a particular quality. (Dowling and Harwood, 1986:214)
Let’s break this down into its constituents – *arousal* and *meaning*.

**Arousal**

According to Dowling and Harwood, autonomous nervous system (ANS) arousal is “content-less” in that emotional qualities cannot be read directly from the qualities of the arousal. The ANS controls many biological functions such as heart-rate, breathing, and digestion. The ANS is a fast acting system that provokes us to react first then interpret later. If a stimulus causes our heart rate increase or our hair to stand on end, we then search for a cognitive interpretation of the stimulus. Cognitive interpretation of the arousal is required before a specific emotion can be construed.

Music involves several overlapping schemata (melodic, harmonic, rhythmic, timbral) all operating simultaneously so that subconscious interruptions to expectations are occurring on different levels. An interruption creates arousal which in turn triggers cognitive activity aimed at interpretation of the novel event. Such an interpretation integrates the event into the meaningful pattern of the piece.

**Meaning**

Music conveys meaning in the relationship among signs embodied in a complex semiotic system. Seeger (1977) can be paraphrased as suggesting that words name things, but music does not name. Dowling and Harwood contend that: “If the meaning of music were entirely accessible at explicit verbal levels, we would not actually have to listen to it – we would just talk about it.” (Dowling and Harwood, 1986: 220) They also point out that “the schematic processes whose interruption causes arousal are mostly subconscious.” (Dowling and Harwood, 1986: 219)

In summary we can say that when listening to music, interruption to ongoing schemata causes arousal. The arousal triggers cognitive interpretation, or search for meaning. Together the arousal and the interpretation may produce an emotional experience. The schematic processes whose interruption causes arousal are mostly subconscious. Musical meaning is mostly non-verbal.
**Arousal, complexity, and stylistic change**

According to Berlyne (1971), the listener seeks arousal, but only up to a point. Beyond that point, the listener seeks to avoid further arousal.

Thus the listener’s preference in relation to musical complexity will follow an inverted U-shaped function (See Figure 4.1).

![Preference vs Complexity Chart](image)

**Figure 4.1: Plot of listener preference vs musical complexity**

Vitz (1966) found that listeners with more musical training and interest preferred higher levels of complexity (dashed curve). He even found that certain mid-range frequencies (400-750 Hz) and intensities were preferred over lower and higher values, presumably because more and more extreme values produce higher and higher levels of arousal. (Vitz, 1972)

There is also a relationship between a particular musical style and its perceived complexity. “As listeners’ schemata accommodate to a style – that is, as they include more and more of its novelties, and subtle nuances - the composer finds it more and more difficult to produce the schematic interruptions necessary for emotional arousal. The composer needs to go outside the existing style for interruptions, and so the style changes.” (Dowling and Harwood, 1986: 224)

Returning now to questions that are directly relevant to the current study: What are the mechanisms for arousal and cognitive interpretation, and how do such mechanisms relate to the musical materials as they unfold throughout the work?

Interpretation is dependent on arousal, so the basic units of acousmatic music must be re-examined, along with their perception, and the nature of relationships between those basic units.
4.3 Recognition of signs and symbols in acousmatic music

In tonal music, it would appear that for Bigand (1993), the basic musical unit and sound unit is the “note event”, whereas in acousmatic music it is the “sound event”. This is the “sign” in semiotic terms. In acousmatic music, disassociating or isolating signs from each other is quite a difficult task. As we saw earlier, many types of discourse unfold simultaneously in a dynamic environment. The fundamental discourses in acousmatic music I shall call source-cause discourse and typological-relational discourse. The former is characterised as semantic discourse, the latter as syntactic discourse.

In the recognition of “signs” one must take account of the multiple interpretation possibilities afforded by these multiple discourses. Firstly, the topic of sound event and sound source recognition will be explored. Then the relationships between identified sound events and the listener will be examined.

4.3.1 Recognition of sound sources and events

In an earlier chapter, McAdams’ stages in the recognition of sound sources and events were observed (McAdams, 1993). After sensory transduction and auditory grouping, the third stage involves the analysis of auditory properties and features. McAdams focuses on the micro-temporal and macro-temporal properties of sound events to suggest that each of these categories reveals its own set of features which carry certain implications (see Table 4-3).
**Table 4-3: Implications of micro-temporal and macro-temporal properties**

Auditory property analysis provides a group of abstract properties which act as input to the matching process with representations in memory. Whether through “comparison” or “activation”, the result of the matching process will either be a match with a category, no match with a category, or too many matches. In the latter two cases, no match will take place.

Categorical matching or lexical activation give rise to associated knowledge of properties of the sound “class” including its relationship to the listener and the local environment: “These associations would allow the listener to plan appropriate action
without having to verbalise what was heard.” (McAdams, 1993:154) Categorical matching and activation of associations establishes a “sense” of the sound.

After activation of a representation or category in memory has taken place, further activation of the listener’s lexicon of names, concepts and meanings associated with the “class” of sound events may occur: “Once language is available, however, the recognition process also gives access to a verbal lexicon that allows the listener to name or describe the event verbally.” (McAdams, 1993: 154-155)

So in relation to sound events and their recognition, the following important terms can be summarized:

**Recognition**

- A sound event matches a representation in memory of a class of similar sounds.

- There is activation of associated concepts and meanings that relate the perceiver to the sound class and the environment. (Note the introduction of the term “environment”.)

- Associations may be non-verbal. For example the sound of a loved-one’s voice may trigger their visual image.

- A sound class’ associated concepts and meanings will have been codified, that is learned, through immersion in, and through interaction with, the environment. The recognition of environmental sounds will be considered further below.

**Identification**

- Identification is where the listener can name or describe the event verbally. For example: “A trotting horse.”

- Identification within the verbal lexicon gives rise to further semantic associations beyond the auditory processing realm.
Partial Recognition

- Partial recognition may take place where an exact match with a particular instance of sound within a sound class does not take place, but where a sound event may be matched with, or judged similar to, the sound class. For example: “bell-like”, “bird-like”, etc.

Non-recognition

- Non-recognition occurs when no match is found. Ambiguity or confusion may result.

The use of recorded “real-world” or environmental sounds is an important contribution to acousmatic music, so attention must now be directed at work that has been carried out on the recognition and classification of environmental sounds.

4.3.2 Identification of environmental sounds

In his PhD thesis, Gygi (2001) asks the question: Why study environmental sound perception?

In attempting to answer this question, Gygi begins by citing research that supports the claim that speech perception and environmental sound perception are separable in the brain (Peretz, 1993), along with a similar disassociation between the production of speech and music (especially the phenomenon of amusia). Gygi argues that environmental sound perception is as specialized as, and distinct from, both speech perception and music perception. Therefore environmental sound perception merits its own investigation.

Gygi claims that the emphasis in listening to environmental sounds is on identifying the source of the sound rather than on the semantic content (speech) or expressive value (music):
The emphasis on identifying the source of a sound as opposed to the semantic or expressive value of the sound itself may result in different ways of listening to environmental sounds, as opposed to speech or music. Certainly, experimental results such as those of Remez et al. show that how we listen to sounds is partially a function of what we expect the sound to be. If the goal is to identify the source of a particular sound, the listener may have to focus on short-term spectral-temporal properties (to identify it as rapidly as possible) as opposed to tracking over an extended period of time, which is required to extract the ‘message’ conveyed by speech or music. Since the variety of sources in the environment is not as limited as in speech or music (see Table I-1), the listener cannot direct attention to a particular bandwidth or modulation rate, and there are far fewer restrictions on ‘lawful’ sequences of sounds. As a result the listener, not having many obvious constraints to help focus listening, would have to learn some system of organization of the sounds. (Gygi, 2001: 5-6)

4.3.3 Research on the perception of environmental sounds

Research on the perception of environmental sounds has been quite meagre when compared with research in other areas of cognitive psychology. Presented here is a summary of the main areas of prior research relevant to the present study. Much of this material is documented in Gygi’s thesis.

Quality ratings and semantic differential - Quality rating studies collected ratings of environmental sounds on scales such as pleasantness and unpleasantness. They aimed to find the acoustic factors that predicted these judgments.

Studies using the semantic differential methodology (Osgood, 1952) asked subjects to rate sounds on category scales using opposite verbal labels (e.g. Heavy-light). The results are subjected to factor analysis to reduce the dimensionality to a few independent scales which have the maximum predictive power.

The problem with these studies is that a lot of the variance in rating is unaccounted for: “In general, clear and definite links between evaluative judgments and stimulus properties have been elusive.” (Gygi, 2001: 21)

Cognitively-oriented studies - Cognitive studies attempted to examine how environmental sounds compared to speech sounds on various cognitive tasks, such as recognition memory, effects of priming on identification, and interference caused by speech or cross-modal stimuli:
In general, the cognitive studies seem to indicate that environmental sounds are remembered differently and less well than speech. Whereas speech, and to a lesser extent, music, can be abstracted away from the auditory stimulus retaining largely the semantic content (with acoustic details saved implicitly in what Pisoni (1993) has called indexical memory), it seems that memory for environmental sounds is more explicitly bound to the details of the waveform. (Gygi, 2001: 24)

**Ecological acoustics and event perception** – Vanderveer (1979) attempted to lay the groundwork for “ecological acoustics” - an extension of the theories of James Gibson. She conducted both identification and similarity studies.

In the identification study, listeners were presented with 30 environmental sounds and instructed to “write down what you hear”. Responses were then analysed for accuracy of identification and for the nature of their descriptions. Identification of sounds was fairly accurate. Description tended to specify the action involved in producing the auditory event, rather than describing acoustic properties or qualitative judgements.

Vanderveer formed a tentative hypothesis “…that information about the actions involved in an acoustic event is largely carried in the temporal patterning, while information about the objects involved in the event is conveyed by the frequency information.” (Gygi, 2001: 26)

In Vanderveer’s similarity study, listeners had to group two sets of 20 sounds into groups based on the similarity of sound. Similarity matrices were derived and analysed qualitatively. Vanderveer concluded that the most salient acoustic determinants of similarity were:

- Temporal patterning variables: Percussive vs. continuous sounds, rhythmic patterning, and attack and decay time.
- Resonance or other characteristics of particular objects, surfaces, and substances, such as metal, paper, and things with a rough texture.

(Gygi, 2001:27)

Vanderveer also noted a clustering based on the type of event that produced the sound, eg. Whistling and blowing sounds grouped together.
Following Vanderveer there were more studies of environmental sounds from an ecological event-perception perspective. Bregman’s huge body of work on auditory streaming and auditory scene analysis can be viewed as a contribution to the ecological acoustic approach.

Warren and Verbrugge (1984) developed a ‘source-filter’ model of auditory event perception. What they called ‘structural invariants’ indicated the spectral features, which in turn denoted the type of material involved in the interaction (eg. Glass bottles). The ‘transformational invariants’ were determined by the temporal patterning and they specified the type of interaction or event (eg. bouncing or breaking).

Richards (1988) developed a model of an acoustic event as an interaction between power sources, oscillators, resonators, and couplers. According to Richards, the sound event conveys information about the source structure. Richard’s theory theory was empirically demonstrated in subsequent experiments.

Freed (1990) looked for acoustic variables that would enable listeners to predict the hardness of a mallet striking a metal plate. Recordings were played to listeners who rated the hardness of mallets on a scale of 1 to 9. Freed subsequently showed that there are measurable acoustic features that predict properties of the source object – “mallets”.

To identify a range of environmental sounds, and not just memorise single instances of a sound, listeners must have some way of coding and organising sounds and sound categories.

Gaver (1993a) created a classification system based on an ecological account of listening:

Inspired by Gibson, he drew a sharp distinction between ‘musical listening’, which is focusing on the attributes of the “sound itself”, and ‘everyday listening’ in which “the perceptual dimensions and attributes of concern correspond to those of the sound-producing event and its environment, not to those of the sound itself.” (Gygi, 2001: 47-48)

Gaver started with the physics of sound-producing events based on three classes: vibrating bodies, changes to liquids, and aerodynamic causes.
Vibrating bodies produce sound as a result of interaction. The resultant sound is dependent on the nature of the objects involved and the type of interaction. Gaver grouped what he called material and configuration by their effects in the temporal or frequency domain. Gaver’s grouping confirms previous assertions that “attributes of the object affect the frequency content of a sound, whereas attributes of the interaction affect the temporal structure”. (Gygi, 2001: 49)

Gaver produced a hierarchy of sound events based on acoustical attributes and listener descriptions. Combinations of the basic sources produce more complex events: “patterned sources (repetition of a basic event), complex sources (more than one sort of basic level event) and hybrid sources (involving more than one basic sort of material). Combining these complex sounds, Gaver developed a ‘map’ of environmental sounds”. (Gygi, 2001: 50)

There are some omissions from Gaver’s scheme including electrical sounds, fire sounds, and speech or animal utterances. Gaver (1993b) also used synthesized sound in order to discover the invariant attributes which characterise certain classes of sound events.

The basic hypothesis of all the ecological based approaches is that “the acoustic signal is informative about the physical events which generated it.” (Gygi, 2001: 53)

**Semantic-syntactic studies** – Another area of research into environmental sounds stemmed from the use of semantic rating scales. Howard and Ballas (1980) used speech research on the interaction between knowledge-driven (top-down) processes and data-driven (bottom-up) processes. They wanted to know whether the same considerations applied for non-speech sounds using an artificial grammar.

The importance of the work of Howard and Ballas in their approach to treating environmental sound as language is such that an extensive discussion of their experiments will be developed in its own section below.

**Summary of prior research** – Gygi quotes the following factors that emerge from the research on environmental sound perception:
• Listeners are very good at rapidly identifying a large number of environmental sounds, as well as specific physical properties of the source of the sound.

• There are large differences across the sounds in both identification accuracy and rapidity. Some of this is due to acoustic factors (properties of the waveform), some due to more central factors (learning, context, causal uncertainty).

• The acoustic and cognitive factors are not wholly separable. Many semantic judgments about these sounds are closely tied to acoustic properties. Further, there seems to be enough information in the waveform to allow judgments about a wide variety of physical attributes of the source of an auditory event, which can aid in identification.

• Because these sounds have meaning, they seem to have some sort of psychological structure, if not exactly a grammar and syntax, which also affects identification. (Gygi, 2001: 70)

4.3.3 Gygi’s Definitions and Experiments

Gygi’s work on the perception of environmental sounds is so comprehensive that several of his most relevant experiments will be detailed here, but first some of his definitions need to be elaborated on.

Gygi (2001) constructs a useful table of comparisons between speech, music and environmental sounds with regard to their basic units, sound sources, spectral structure, temporal structure, syntactic and semantic structures (See Table 4-4).
<table>
<thead>
<tr>
<th>Units of Analysis</th>
<th>Possible Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Speech</strong></td>
<td><strong>Music</strong></td>
</tr>
<tr>
<td>Phonemes</td>
<td>Notes</td>
</tr>
<tr>
<td>Finite: Human vocal tract</td>
<td>Finite: Instruments (so crafted or designated)</td>
</tr>
<tr>
<td><strong>Environmental Sounds</strong></td>
<td><strong>Events</strong></td>
</tr>
<tr>
<td></td>
<td>Possibly Infinite: any naturally-occurring sound-producing event</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Spectral Structure</th>
<th>Temporal Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Speech</strong></td>
<td><strong>Music</strong></td>
</tr>
<tr>
<td>Largely harmonic (vowels, voiced consonants); tend to group in formants. Some inharmonic (stops, fricatives)</td>
<td>Largely harmonic, some inharmonic (percussion)</td>
</tr>
<tr>
<td>Short (40 ms - 200 ms). More steadystate than dynamic. Timing constrained but variable. Amplitude modulation rate for sentences is slow (~ 4 Hz)</td>
<td>Longer: 600-1200 ms (Fraisse, 1956)) Mix of steady-state (strings, winds), and transient (percussion). Strong periodicity</td>
</tr>
<tr>
<td><strong>Environmental Sounds</strong></td>
<td></td>
</tr>
<tr>
<td>Proportion of harmonic vs inharmonic not known, although both exist</td>
<td>Longer (500 ms - 3s). Proportion of steady-state to dynamic not known. Variable periodicity</td>
</tr>
</tbody>
</table>

**Table 4-4: Structural features of speech, music and environmental sounds.**

(Adapted from Gygi, 2001: 9)
Gygi creates the following definitions for the purpose of his study:

**Environmental sounds** - Naturally occurring, non-speech, non-music sounds.

**Sound sources** - The object or objects from which sounds emanate.

**Events** - Things done by or to sources that yield propagated acoustic waveforms.

Events have closely associated sounds and this close association enables a listener to identify a source or an event. Environmental sounds have relevance to listeners and this “meaning” involves higher-level cognitive factors, in addition to acoustic features, involved in their identification.

There is quite a bit of covariance between the acoustic and semantic factors: “Sounds which may indicate, say, a tiger or lion approaching its prey would tend to be similar, because they would be produced by similar sources (large creatures) engaged in similar activities (stalking through foliage).” (Gygi, 2001: 8)

Certain acoustic properties, such as first and second formants, have predicted identification of a class of speech sounds – vowels. Grey (1977) identified attack and spectral envelope as determinants of music instrument timbre. Analogous features that allow the identification of environmental sounds have been very hard to find. The goal of Gygi’s research is to isolate the acoustic factors that are predictive of identification of environmental sounds.

Gygi’s study reports on:

1. Identification of band-pass filtered environmental sounds.

2. Identification of event-modulated noise.

3. A determination of the main determinants of identifiability: simple acoustic features (eg. Power, duration, spectral centroid) or higher-order spectral-temporal attributes.
In the following discussion, the details and findings of Gygi’s experiments concerning the perception of environmental sounds have been extracted for only those experiments and findings that are relevant to the present study.

**Similarity studies**

From all of the types of studies of environmental sounds, the most relevant to our investigations are the “similarity studies”.

In listening to acousmatic music, or analysing acousmatic music, events occurring at one point in time are compared with events occurring within the next time segment. An assessment of repetition, variation, or contrast is made. Such comparisons involve a test of similarity of events separated in time.

Gygi (2001) presents results of similarity studies conducted at the *Hearing and Communication Laboratory* at Indiana University. He points out that for “meaningful” sounds, there needs to be a consideration of their “psychological space”.

One experiment attempted to uncover the structure of such a psychological space by using multidimensional scaling (MDS) procedures in order to discover what dimensions describe such a space. The technique has also been employed by Grey and Moorer (1977) for musical instruments and Lakatos et al (1997) for acoustic correlates of source features.

Gygi conducted three experiments using 50 environmental sounds. In the first “acoustic similarity” experiment, recordings of 100 instances of the environmental sounds were played to listeners who were asked to rate the similarity of each pair on a seven point scale from least similar to most similar.

The second experiment tested “sound image similarity” involving only “imagined sound”. Instead of presenting the participants with sound recordings, they were given a pair of labels for the 50 environmental sounds and asked to: “Make similarity judgements based on your memory or knowledge of these sounds.” (Gygi, 2001: 87)

The third experiment, which Gygi characterised as “source image similarity”, involved an “imagined event”. Participants used the same set of paired labels and were
instructed to: “Make similarity judgements based on how similar these events are compared to all other events in the list.” (Gygi, 2001: 87) In this experiment, sound was not even mentioned.

A two-dimensional MDS solution was derived for each experiment. An interesting result was that the three solutions were very similar and all resembled the results for the “acoustic similarity” study (See Figure 4.2)

![ACOUSTIC SIMILARITY](image.png)

**Figure 4.2: The 2-D MDS solution for Gygi’s “acoustic similarity” experiment.**

Reproduced from (Gygi, 2001: 89)

Similar clusterings are found in each plot: “…animal sounds, more rhythmic sounds (clapping, ping pong ball bouncing), and water-based sounds tend to cluster.” (Gygi, 2001: 88). Interestingly the sounds of animal and human utterance not only group with each other, they also cluster with sounds of signification and warning: whistle, car honk, phone.
Gygi searched for a meaningful interpretation of the two dimensions: “If the dimensions are all based on semantic content, one would not expect to find acoustic variables that would account for the ordering on either of the dimensions in a meaningful way.” (Gygi, 2001: 88)

Gygi measured over twenty acoustic variables covering different aspects of spectral distribution: energy in frequency bands, amplitude envelope, spectral centroid velocity, autocorrelation, pitch measures, cross-channel correlation, spectral similarity.

The values of all these variables were correlated with the ordering of the sounds on both Dimensions 1 and 2. Based on the clusterings, it was hypothesized that some measures of ‘pitchiness’ or harmonicity might account for the ordering on the first dimension. Dimension 2 is not quite so straightforward, but it was thought that some measure of rhythmicity might account for the variance on Dimension 2. (Gygi, 2001: 88-89)

The highest correlations for Dimension 1 were: Spectral spread; and confidence in the pitch of the signal (see Slaney, 1994).

Correlations with the second dimension were not so clear cut. There were correlations with some variables, but the correlations were low. These low correlations were with two main types of variables: Measures of rhythmicity and continuity; and measures indicating high frequency energy. The most significant correlations were to do with measures of rhythmicity and spectral mean.

Based on these similarity studies, Gygi’s overall conclusions regarding the psychological space for environmental sounds were that it is the same when listening to sounds and thinking about sounds, and:

This space seems to have some basis in the acoustic properties of the sounds, although it is likely complicated by other, perhaps semantic, factors. Certainly the underlying dimensions are not nearly as clear cut as for laboratory-generated complex sounds (Howard & Silverman, 1976) or for narrowly defined sets of environmental sounds (Howard, 1977; Cermak & Cornillon, 1976).

Gygi’s subsequent experiments seemed to explore the two identified dimensions with experiments of their own. They were spectrum-related experiments and event-related experiments.
Identification of filtered environmental sounds

Gygi subsequently conducted an experiment on the identification of bandpass-filtered environmental sounds in order to test spectrum-related identification considerations.

From the bandpass-filtered sound experiments Gygi found that the discrimination of environmental sound is quite robust to the effects of filtering, and the necessary information for identification is in the middle and upper frequencies (600-4800 Hz). Sounds can be grouped into similar identification profiles. Gygi created 8 clusters from 70 environmental sounds.

Although there was a spectrum-related explanation for the grouping of a number of clusters, finding a spectral basis for several clusters was problematic. Gygi hypothesised that some identification results could be better explained by an event-based theory “in which sounds generated by similar events tend to have similar identification profiles.” (Gygi, 2001: 109) For example, sounds in “Cluster 4” could be could be grouped in five event-type classes (See Table 4-5).

<table>
<thead>
<tr>
<th>Vocal Tract</th>
<th>Signaling</th>
<th>Mechanical</th>
<th>Water-based</th>
<th>Friction-Based?</th>
</tr>
</thead>
<tbody>
<tr>
<td>BABY</td>
<td>SIREN</td>
<td>CAR STARTING</td>
<td>POURING</td>
<td>ROCKING</td>
</tr>
<tr>
<td>DOG</td>
<td>HONKING</td>
<td>HELICOPTER</td>
<td>GARGLE</td>
<td>SCREEN DOOR</td>
</tr>
<tr>
<td>CAT</td>
<td>BELLS</td>
<td>ELECTRIC SAW</td>
<td></td>
<td>CLOCK</td>
</tr>
<tr>
<td>COW</td>
<td>CYMBAL</td>
<td>PRINTER</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ROOSTER</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COUGH</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4-5: Grouping of sounds into five classes according to event type.

Reproduced from Gygi (2001: 109)

Consideration of an event-type basis led to the next experiment.

The Development of a Cognitive Framework for the Analysis of Acousmatic Music 83
Identification of environmental sounds from event-modulated noise

This experiment consisted of identification of environmental sounds from event-modulated noise. The idea is to eliminate spectral information and retain temporal information.

Gygi’s event-modulated noise studies showed that:

... as with speech, in the absence of frequency information temporal information is sufficient to identify many (with these stimuli, 50% or 35 out of 70) environmental sounds. The sounds which tend to be identified best under these conditions are sounds with some temporal patterning, and there is a strong effect of experience with the sounds, presumably because listeners learn the temporal features important for identification. (Gygi, 2001: 126)

Acoustic factors in EMN identification

In the next set of experiments, Gygi attempted to isolate the specific acoustic factors that accounted for the event-modulated recognition results, that is the specific temporal recognition factors.

He concluded that there are at least five variables required to represent what listeners perceived about temporal structure, and there may be higher order variables that are a combination of several of the lower order set. If some spectral information is added back in, listeners will change the focus of their auditory attention.

Gygi’s environmental sound similarity and identification studies seemed to indicate that one should really consider sub-classes of sounds and he postulated possible bases for classification:

- Acoustic features, such as the harmonic/inharmonic continuum which was the major determinant of similarity; the frequency region which allows the best identifiability; or, perhaps on the salience of temporal structure.

- The sources and events which produced the sound, as Gaver (1993a,b) detailed. So researchers would differentiate between ‘impact’, ‘water-based’, ‘air-based’ and ‘simple’, ‘hybrid’ and ‘complex’ environmental sounds.
Higher order semantic features, such as causal ambiguity, ecological frequency or importance to the listener. Examples: “alerting sounds,” “rewarding sounds,” “ignorable sounds.”

Some unexplained scheme, such as the identification of profile clusterings. (Gygi, 2001:155-156)

Of these bases, Gygi considers the acoustic features as of prime importance (accounting for 60% of the variance in identification certainty), however he acknowledges that multiple levels of processing operate simultaneously from the bottom-up (peripherally) and top-down (centrally). He utilises the analogy with speech where phonology, syntax, and semantics all contribute to perception of speech sounds.

4.3.4 Factors in the identification of environmental sounds

Overall Gygi stated that his results show:

- The most important frequencies for identification of environmental sounds are in the 1200-2400 Hz region, comparable to that found for speech. However, environmental sounds seem to have a greater amount of information in frequencies above 2400 Hz than speech.

- Environmental sounds can be severely high-pass and low-pass filtered and still be identifiable, even more so than speech.

- As with speech, the ability to recognize environmental sounds is extremely robust despite removal of spectral information. Like speech, environmental sounds have a multiplicity of acoustic cues available to listeners due to their complex spectral-temporal composition. This redundancy aids in the robustness to signal degradation.

- A combination of central and peripheral factors are involved in the identification of environmental sounds, which is also a hallmark of speech perception.

- Although the issue of whether ‘speech is special’ has not been resolved, it does seem that environmental sounds are much more ‘speech-like’ than complex
laboratory-generated sounds. This may be largely due to the even greater complexity of environmental sounds, which vary along multiple dimensions in both the time and frequency domains. However, the fact that environmental sounds have meaning for the listener undoubtedly plays a part. Future studies should continue to investigate the acoustic determinants of environmental sound perception, but should also consider the interaction of the cognitive and sensory factors. (Gygi, 2001: 157-158)

Gygi’s identification studies provide a basis for isolating features of environmental sounds that could be singled out for analysis. The events which produce the sound could also be examined, as Gaver (1993a, 1993b) detailed. Another important factor is the classification notion of ‘simple’, ‘hybrid’ and ‘complex’ environmental sounds.

Important questions remain: What of the top-down processes such as semantic features? Are there syntactic relationships between environmental sounds, and if so, what are the effects?

To answer the first question with regard to semantic features, the work of Howard and Ballis will be examined, and then the implication-realization model of Narmour will be presented as a practical example of the operation of relationships in a musical context.

4.4 Syntactic and semantic factors

In an article by Howard and Ballas (1980), they documented their experiments which tested syntactic and semantic factors in the classification of non-speech transient patterns. What is important about this series of experiments is that they are concerned with patterns of sounds (i.e. sequences of sounds) – both pure tones and environmental sounds.

Experiment one tested the influence of syntactically structured sequences in the identification of a pattern (a sequence) of simple tones.

Experiment two examined the effect of syntax in a sequence of environmental sounds. These were more difficult for listeners to classify than simple tones. There was quite a significant difference from experiment one. The differences noted between the use
of simple tones and environmental sounds led Howard and Ballas to devise a third experiment that tested the influence of semantic factors when recognisable, non-speech environmental sounds are used.

The importance of their work to the present discussion is such that their experimental design, results, and findings will be summarized here.

4.4.1 Background to the work of Howard and Ballas

Howard and Ballas begin by noting that speech recognition research has demonstrated both top-down (knowledge driven) and bottom-up (data driven) processes. When listening to speech, “listeners appear to use their general knowledge of linguistic structure (both syntactic and semantic), as well as specific information in the signal.” (Howard and Ballas, 1980: 431)

There has been relatively little research into semantic and syntactic factors in the perception of non-speech sound patterns. Howard and Ballas note that there are many everyday sound patterns with an inherent syntax in the form of an ordered sequence of what they call “transients”. Such “transient patterns” are typified in the sound effects for radio drama for example.

Their paper attempts to shed some light on syntactic (temporal structure) and semantic (knowledge of source events) factors in transient pattern recognition. The authors cite top-down processing influences from the speech processing literature, such as the phenomenon of “learning” missing or mispronounced words in a passage of speech, as found by Warren (1970) and others.

With regard to non-speech perception, the work of Bregman (1978) is quoted as demonstrating the role of knowledge-driven (top-down) processes in the perception of simple tonal sequences. They remind us of Bregman’s assertion that we pay attention to acoustic sources, not acoustic components. Audio streaming is used to sort the perceived acoustic world into separate sources. Bregman concludes there are two sets of heuristics (or rules) which may apply to auditory streaming, firstly a set to parse all the signals, and a second set “related to a listener’s skills, intentions, and knowledge of the stimuli that apply only to selected patterns.” (Howard and Ballas, 1980: 432)
Bregman’s work concentrates on the first type of heuristics whereas the work of Howard and Ballas addresses the second type.

The approach of Howard and Ballas draws on Reber’s “implicit learning” research which examined syntactic structure’s role in the classification of letter patterns (Reber, 1969). A finite state grammar was used to generate grammatically consistent letter patterns and Reber found “superior classification performance with grammatical, as opposed to non-grammatical, arbitrarily grouped letter patterns.” (Howard and Ballas, 1980: 433)

The finite state grammar is reproduced in Figure 4.3.

![Finite State Grammar](image)

**Figure 4.3: State-transition diagram used by Howard and Ballas.**

Reproduced from Howard and Ballas (1980: 432)

Grammatically correct patterns are generated by following the diagram from its input (on the left) to its output (on the right). A circular arrow in the state-transition diagram indicates that a letter can be repeated any number of times. A legal pattern would be “AAACDD”, whereas “AADDCC” would be considered ungrammatical.
Howard and Ballas employ the same finite state grammar in each of three experiments to generate “grammatical” sequences of firstly simple tones, and then complex environmental sounds, which they characterise as syntactically structured patterns:

In the first experiment, listeners classified meaningless patterns of brief-duration pure tones. The second experiment was similar to the first, but the pattern components consisted of complex familiar sounds rather than simple tones. …Finally the role of both syntactic and semantic factors was investigated in the third experiment, in which some listeners classified semantically interpretable patterns of complex sounds. (Howard and Ballas, 1980: 433)

4.4.2 Experiment one – Syntactic factors with simple tones

Experiment one was designed to test whether listeners can use syntactic information to assist the classification of non-speech sound patterns consisting of a short sequence of pure tones (1157 Hz, 1250, 1345, 1442, and 1542 Hz) arranged in varying sequences according to pitch. Duration and loudness were held constant.

Listeners were divided into two groups: grammatical and non-grammatical. Listeners were told they would be hearing patterns, and some of the patterns were “targets”. Their task was to pick out targets.

The grammatical group’s targets were generated by the finite state grammar. For the non-grammatical group, the targets were randomly generated (and non-grammatical). Both groups were presented with 12 target patterns (repeated 4 times) plus 48 non-targets.

In a final trial block, listeners were first told a set of rules were used to create target patterns. They then completed 96 more pattern trials but without feedback. This test was included to see whether listeners in the grammatical group could use their syntactic knowledge to classify new grammatically-generated patterns.

Several of the more interesting conclusions from this experiment were:

- “The grammatical group reached a substantially higher overall performance level than did the non-grammatical group.” (Howard and Ballas, 1980: 433-434), and
“Overall, performance on the final block supports our position that listeners in the grammatical group had actually learned something about the syntactic rules used to generate the target patterns they had classified previously.” (Howard and Ballas, 1980: 434)

4.4.2 Experiment two – Syntactic factors with environmental sounds

In Experiment one, Howard and Ballas had used patterns consisting of a sequence of simple tones. Experiment two was designed to test whether the same results would be found for patterns made up of complex, realistic sounds.

The methodology of experiment two was the same as for experiment one except that the individual sound events sequenced together to form patterns were familiar, but unrelated, real-world sounds: drill, clap, steam, clank, wood.

From the results of experiment two, Howard and Ballas were able to make similar conclusions: “that listeners are able to use syntactic or grammatical structure to facilitate classification in both cases.” (Howard and Ballas, 1980: 435)

The results also showed that the performance of the grammatical group in experiment two, using complex sounds, was considerably lower than for the grammatical group in experiment one, which used simple tones.

Howard and Ballas attempted to explain the differences between these two experiments by referring to semantic processing:

Since the sounds were familiar, the listener could not avoid using a parsing strategy that tried to make “sense” out of the patterns. Since the finite state grammar we used was semantically arbitrary, this could prove an impossible task. (Howard and Ballas, 1980: 435)

The final test block with novel grammatical patterns demonstrated poor performance for both groups – grammatical and non-grammatical. Was the semantic processing interfering in some manner with the classification task? A third experiment was devised to test the influence of any semantic factors.
4.4.3 Experiment three – Syntactic and semantic factors with environmental sounds

Experiment three was designed to evaluate the semantic component in auditory classification. In this experiment Howard and Ballas used real-world sounds to create patterns, but some of these sounds were required to be semantically sensible. Thus some patterns had to be both syntactically and semantically reasonable through being a recognisably plausible sequence of related sounds.

The same finite state grammar was used to produce a series of water and steam related events that also made semantic sense in the sequence. For example, the grammatically correct output string AAACDD corresponds to opening a valve with three turns, followed by a single release of steam, which causes the pipe to clang twice.

Four groups were tested in this experiment with two groups receiving a suggestive statement prior to testing (semantic priming).

The groups were:

1. Grammatical (syntactic variable) with semantic instructions (semantic variable).
2. Non-grammatical with semantic instructions.
3. Grammatical with no semantic instructions.
4. Non-grammatical with no semantic instructions.

The source sounds used were: valve, drop, steam, clang, flush.

Other than the semantic instructions and the different source sounds, the methodology was the same as for experiment two.

From the results of experiment three, Howard and Ballas concluded the following:

- The grammatical groups performed at significantly higher level than the non-grammatical groups (approaching the significance of experiment one).
• The grammatical/semantic group performed significantly better than the grammatical/non-semantic group.

• In the no-feedback final test block, both grammatical groups performed at above chance levels, whereas both non-grammatical groups performed at approximately chance levels. Thus grammatical group listeners were able to internalise aspects of the pattern’s grammar regardless of semantic instructions.

So if acoustic, semantic, and syntactic factors are operating in the identification of sequences of environmental sounds, what are the relative contributions provided by each of these factors?

4.4.4 Contribution to identification by acoustic, semantic, syntactic and other factors

In 1993, James Ballas published an account of five experiments under the title *Common factors in the identification of an assortment of brief everyday sounds* (Ballas, 1993). Acoustic, ecological, perceptual and cognitive factors that are common in the identification of 41 brief, varied sounds were evaluated. In Experiment One, identification time and accuracy, causal uncertainty values, and spectral and temporal properties of the sounds were obtained. Experiment Two was a survey to obtain ecological frequency counts. Experiment Three solicited the perceptual-cognitive ratings. Factor analyses of spectral parameters and perceptual-cognitive ratings were performed. Identification time and causal uncertainty are highly interrelated, and both are related to ecological frequency and the presence of harmonics and similar spectral bursts. Experiments Four and Five used a priming paradigm to verify relationships between identification time and causal uncertainty and to assess the effect of sound typicality. His results supported a hybrid approach for theories of everyday sound identification.

**Experiment One: Identification time and causal uncertainty**

The first experiment was conducted to determine the causal uncertainty values and identification response times for a set of sounds. The sound set did not include animal vocalizations. Listeners were asked to press the space bar to initiate a sound and to
press it again as soon as they had an idea about the cause of the sound. Response times were measured. Listeners also had to type an identification of the sound, in the form of both a noun and a verb. Listeners were presented with 41 test sounds.

In a second part of the experiment, listeners were presented with the 41 sounds again, and were able to provide an alternative to the original response, if they wished. The participants then completed a questionnaire designed to assess their familiarity with the events that had produced the sounds.

From the typed responses, a measure of causal uncertainty (Hcu) was determined for each sound. Response times were averaged so that each sound was given a mean identification time (MIT). A correlation was found between the median causal uncertainty values and the mean identification times. The relationship appeared to be non-linear, and a stronger correlation was found with the log of the mean identification time (LMIT).

Examination of the causal event familiarity ratings and their correlation with identification time and causal uncertainty suggested that event familiarity is involved in everyday sound identification performance. This issue was followed up in Experiment Two.

**Acoustic factors in identification time and accuracy**

The 41 sounds used in Experiment One were also subjected to acoustic analyses in order to determine what acoustic properties may influence performance time and accuracy. Average acoustic properties included: duration, average magnitude, peak magnitude, power, Fast Fourier Transform spectrum (FFT), 1/3 octave spectrum – Seventeen bands with centre frequencies from 200 Hz to 8000 Hz.

In Experiment One, the mean identification time correlated inversely with (in decreasing order of strength):

- The presence of harmonics
- The presence of similar spectral patterns in bursts
- The presence of continuous bands in the sound
• The maximum FFT component in the spectrum
• The ratio of burst duration to total duration (positive correlation)

Accuracy of identification correlated with:

• The presence of harmonics
• The presence of continuous bands in the sound
• The kurtosis of the FFT distribution
• Spectral components – middle to high bands (1000-2000 Hz)
• The ratio of burst duration to total duration
• The number of bursts
• The presence of similar patterns in bursts
• The duration

Better results were obtained with combinations of spectral and temporal variables. The union of harmonics in continuous sounds and similar spectral bursts in non-continuous sounds (HB) produced improved correlations with mean identification time and accuracy. The addition of the ratio of burst duration to total duration and spectral skewness factors further improved LMIT correlation.

Most significantly, acoustic factors, when combined, can account for up to 50% of the variance in identification time.

What other factors are involved? Experiment Two examined the role of ecological frequency and Experiment Three examined other perceived properties not directly related to acoustic factors, such as cognitive knowledge about the sounds.

**Experiment Two: Ecological frequency study**

In Experiment One, the familiarity of sounds to the participants was found to correlate with identification time and causal uncertainty in a small but still significant way.
Experiment Two set out to examine whether sound identification performance was related to the occurrence frequency of sounds in the natural world.

Ballas felt that: “It is likely that the natural occurrence frequency of the sounds used in Experiment One may be quite varied, and the issue should be addressed.” (Ballas, 1993: 254)

As there is little data regarding the occurrence frequency of specific sounds, Ballas designed an experiment whereby college students could report sounds that they heard at specific times during the day of the course of about a week.

A total of 1,185 sounds were reported, distributed by the time of day, day of the week, and context (home, work, school, travel, shopping, outdoor setting, street).

Twenty two of the 41 sounds used in Experiment One were present in the students’ survey. Their ecological frequency counts (Ef) were found to correlate significantly with both causal uncertainty (Hcu) and mean identification time (LMIT).

Ballas then employed regression analysis to estimate identification time (LMIT) using combinations of variables from the spectral and temporal sets and the ecological frequency data. The most significant model comprised of: harmonics in continuous sounds and similar spectral bursts in non-continuous sounds (HB), the average frequency of the spectrum, the ratio of burst duration to total duration, and ecological frequency. This combination accounted for 75% of the variance in identification time. Ballas notes that the variables represent four types of data: spectral, temporal, envelope, and ecological. He concluded:

This result means that performance on a varied set of everyday sounds is related best to a set of measures in different domains rather than a single measure on these sounds. (Ballas, 1993: 256)

**Experiment Three: Perceptual and cognitive judgements**

We have seen in the first two experiments that acoustic properties and ecological frequencies can account for 75% of the variance of identification time, but what contribution do perceptual and cognitive judgements make? This was the motivating factor for Experiment Three. An additional motivation was that perceptual and
cognitive judgements may be useful in developing categorizations of everyday sounds that may reveal the structure of cognitive knowledge representations.

Listeners were asked to rate individual sounds on a variety of perceptual and cognitive scales which were derived from other timbre studies and verbal research.

The 22 rating scales used were from these broad categories:

1. Aural properties, eg. Clarity, loudness, timbre.

2. Conditions antecedent to the identification process, eg. Existence of mental stereotype, familiarity of the sound, number of similar sounds.

3. Aspects of the sound identification process, eg. Necessity of envisioning a sound in a sequence to identify it (context independence), ease of thinking of words to describe a sound.

Ballas produced a table of the rating scales - see his Appendix C. (Ballas, 1993)

Working with a computer and headphones, participants initiated the sound then rated the sound on each of the scales in turn. They could listen to the sound again at any time. Having provided a response on each of the rating scales they moved on to the next sound. The 41 sounds used in previous experiments were used in this experiment.

An interesting side result was that the ease of identifying the cause of a sound was highly correlated with both the ratings assessing the action of the cause and the agent of the cause.

To simplify the participants’ judgements, Ballas performed a principal components analysis to tease out the main factors. Three factors were found to account for over 84% of the variance:

Ratings correlated with identifiability (37%):

- Ease with which a mental picture is formed of the sound.
- Familiarity of the sound.
• Identifiability in isolation (context independence).
• Similarity of the sound with mental stereotype.
• Ease in using words to describe the sounds.
• Clarity of the sounds.

Ratings of sound timbre (31%):
• Relaxed-tense.
• Round-angular.
• Dull-sharp.
• Pleasant-unpleasant.
• Loudness.

Ratings related to the “oddity of the sounds” (16%):
• Ratings of the number of sounds in the same category.
• The number of similar sounds.
• The number of events which could cause the sound.

The issue of ‘categorisation’ was addressed during analysis with Ballas performing a complete linkage cluster analysis. Ballas specified the results in a tree diagram (See Figure 3, Ballas, 1993: 259), but a simpler form is shown below.

There are four main clusters:

1. Water sounds:
   a. Produced with water: drip, splash, bubble, flush.
   b. Water context: boat whistle, foghorn.
   c. Other: clock ticking, lighter (on edge of cluster)
d. Sounds have low sound-timbre scores.

2. Signalling and danger sounds:
   
a. Signalling sounds: telephone, door bell, bugle, sub horn, car horn.
   
b. These have high identifiability and high timbre scores (hard, angular, sharp, tense, unpleasant).
   
c. Sounds connoting danger: fireworks, power saw, automatic rifle.
   
d. These both have low identifiability and are on the edge of the cluster.

3. Door sounds and sounds of modulated noise:
   
a. Door sounds: jail door closing, door opening, electric remote door buzzer, door latched.
   
b. Engine sounds: car backfire, car ignition, lawn mower, tree saw.
   
c. Others: bacon frying, rifle shot outdoors.
   
d. All sounds have low identifiability scores, and variation in timbre.

4. Sounds with transient components:
   
a. Sounds with 2 or 3 transient components: light switch, stapler, footstep, clog step, phone hang, file cabinet, door knock, hammer, cork pop, door close.
   
b. Single transient sounds: tree chop, rifle in-doors.
   
c. Others: two bell sounds, touch tone.
   
d. Most of these sounds have low oddity scores and vary in identifiability.

Ballas went on to reduce the four categories to two: water sounds and impact sounds. He does caution that this outcome may be the result of only three perceptual-cognitive dimensions and a limited sound set: “For example there were not many friction sounds such as sandpapering, tires squealing, and metal grinding.” (Ballas, 1993: 263)
Ballas conducted a further two experiments relating to sound labelling and causal uncertainty. They also explored the issue of priming. However as the experiments are only peripherally related to the current study, they won’t be elaborated here.

**Summary of the Ballas experimental findings**

1. Sound identification is influenced by several factors, including:
   
   a. Acoustic variables.
   
   b. Ecological frequency.
   
   c. Causal uncertainty.
   
   d. Sound typicality.

2. Three quarters of the variance in identification time can be related to acoustic variables and ecological frequency.

3. Combinations of acoustic variables are required to relate to aspects of identification performance.

4. Causal uncertainty is the uncertainty a listener has in determining the possible cause of the sound. Causal uncertainty is strongly related to identification time and accuracy, but only weakly related to ecological frequency.

5. Signal sounds had lower causal uncertainty than that estimated by ecological frequency. Impact sounds and noise modulation sounds had higher causal uncertainty than that predicted by ecological frequency.

6. Sound identifiability is related to:
   
   a. The ease with which a mental picture of the sound is formed.
   
   b. Context independence.
   
   c. The familiarity of the sound.
   
   d. Similarity with mental stereotype.
e. Ease in using verbal description of the sound.

f. Clarity of the sound.

7. Sound clusters were formed on the basis of the factor dimensions:

a. Sound identifiability.

b. Timbre.

c. Oddity.

8. Acoustic variables were found that accounted for half the variance in identification time and accuracy. The best results involved a combination of acoustic variables.

9. Average spectral properties didn’t correlate highly with performance. Confirming other experiments that spectral variables relate weakly to identification.

10. Generalisation beyond the specific 41 sounds should be treated with caution.

11. A hybrid theory is important for everyday sound identification because of the wide variety of everyday sounds.

12. A hybrid theory should incorporate the effect of contextual information.

4.5 Relationship of signs and symbols in acousmatic music

Subsequent to, or simultaneously with, the recognition of sound events and the cognitive activation they induce, we must also consider the relationships between sound events.

The analysis of language processing has shown us that we fuse phonemes into words and chunk (or group) words into phrases and phrases into sentences. Sentences can then be grouped into topics or concepts.

In the act of language processing (spoken or written) we are able to take account of recent sentences, process the incoming speech (or text), and anticipate what may
follow – all in working memory. We are able to revise and update what we’ve recently perceived as well as generate an expectation of what is to follow. Such expectations may be confirmed or denied. We are even able to “fill-the-gaps” when some of the sensory input is partially degraded, for example when one is unsure of a speaker’s accent.

The various forms of grouping with respect to acousmatic music will be considered in the following sections. In processing acousmatic music, we will also find it useful, as Smalley reminds us, to contrast the functional with the incidental and to contemplate the interplay between what I shall call typological-relational functions (syntactic) and semantic functions.

**4.5.1 Simultaneous and sequential grouping of frequencies and sounds**

In listening to acousmatic music then, we can speculate that sounds will be grouped simultaneously and/or sequentially according to some criteria that are both conscious and sub-conscious, influenced by both higher-level cognition and lower-level perception.

Simultaneous grouping of frequencies will generally result in spectral fusion, or what Smalley (1994) calls integration, and will result in some form of single sound event recognition.

Simultaneous grouping of frequencies may be the result of the application of the Gestalt laws so that certain spectral patterns appear to emanate from a single source. The Gestalt law of common fate would suggest that frequency components would be grouped together where it appears they arise from a common cause.

Recall the factors that Bregman highlights as influencing spectral integration (fusion) are:

- Common fate
  - Correlated changes in frequency (micro-modulation, slow modulation)
  - Correlated changes in amplitude (onset and offset)
• Other correlations and correspondences

• “Old-plus-new” heuristic

• Frequency separation

• Partial intensity

• Harmonicity

• Spatial origin

• Textural features (e.g., Irregular spectra)

Separate sounds that each have their own integrated spectra may also be grouped together simultaneously because of the common fate/common cause principles, for example a car brakes suddenly and the squealing tyre sounds are grouped together as belonging to the same car source.

Bregman tells us that sequential integration is the process of temporal grouping of several events, that take place one after the other, into a single auditory stream. Sequential integration uses the changes in the spectrum and the speed of such changes as clues to correct grouping.

The factors identified by Bregman as influencing sequential grouping are:

• Fundamental frequency proximity

• Temporal proximity

• Spectral shape (timbre)

• Intensity (including onset and offset)

• Spatial origin

So the major “data-driven” (or acoustic) factors for sequential integration (or grouping) are similarity and proximity. This has also been confirmed by Gygi’s work on spectral similarity and temporal shaping (Gygi, 2001).
We have also seen, through the work of Howard and Ballas (1980), that there are semantic factors in the recognition (and grouping) of environmental sounds.

Symbolic interpretation of an acousmatic work will therefore be dependent on simultaneous and sequential grouping of possible sources through the evaluation of acoustic, semantic and syntactic factors. We could think of these factors as forming a three dimensional space, upon which recognised sources could be plotted. Each of these dimensions would be multi-dimensional in its own right.

Sequential grouping and processing of sound events is also strongly influenced by listener expectation and attention. Expectation provides the fundamental basis for one model of syntactic processing as applied to tonal music: the implication-realisation model (Narmour, 1989). Does this model have anything to offer our framework for the analysis of acousmatic music?

To explore these discussion themes further, this chapter will conclude with a closer examination of Bregman’s auditory scene analysis ideas, applied to music in particular, and a brief account of Narmour’s implication-realisation model of melodic organisation in tonal music.

4.5.2 Primitive and schema-based segregation compared

What are the relationships between top-down and bottom-up processes that a listener employs when listening to sound and music? In this section, a summary of Bregman’s ideas on what he calls primitive and schema-based organization is presented (Bregman, 1999). This material is presented in an attempt to separate out some of the knowledge-based factors from the perceptual-based influences.

Much of Bregman’s book focuses on primitive processes in scene analysis, that is simultaneous and sequential stream segregation and/or integration, but:

The experiences of the listener are also structured by more refined knowledge of particular classes of signals, such as speech, music, machine noises, and other familiar sounds of the environment. Psychologists argue that this knowledge is captured in units of mental control called schemas. (Bregman, 1999: 665-666)
Each separate schema includes information about a particular regularity in the environment. As noted earlier, Bregman states that regularity can occur in different levels of granularity in size and time. He gives the following examples: A schema for the sound “a”; One for the word “apple”; One for the grammatical structure of a sentence in which the word apple appears; One for the pattern of give and take in a conversation about apples.

Schemas become active when they detect a pattern that they are “programmed” to deal with. When part of the evidence is present, and a schema is activated, it can prepare the perceptual process for the remainder of the pattern. Bregman maintains that this preparation is schema-governed attention.

Schemas can be activated by sense data or other schemas. For example: Reading the word “apple” can prepare for reading the word “fruit”. Schemas do not just fire off in parallel, but rather they compose themselves to create a consistent representation of the world:

(An) example of the schema-based selection of sense data occurs in the phonemic restoration of a speech sound that has been masked by noise (we hear the speech continue through the noise). Apparently we select certain frequency components out of the noise and hear them as if they were the missing speech sounds. This selection must be accomplished by a process that expects certain sounds to be there. Presumably, that process is a schema that represents the sound pattern of a particular word. (Bregman, 1999: 667)

Table 4-6 shows the differences between primitive and schema-based scene analysis.
<table>
<thead>
<tr>
<th><strong>Primitive Segregation</strong></th>
<th><strong>Schema-based Segregation</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>No past learning</td>
<td>Employs knowledge</td>
</tr>
<tr>
<td>No voluntary attention</td>
<td>Uses voluntary attention</td>
</tr>
<tr>
<td>Present in infants. Probably innate</td>
<td>Knowledge must be learned</td>
</tr>
<tr>
<td>Partitions sensory evidence. Partitions into different sound-generating events</td>
<td>Selects from evidence without partitioning it</td>
</tr>
<tr>
<td>Uses relations and clues that apply over wide classes of acoustic events (Innate heuristics)</td>
<td>Uses schemas developed for a particular class of event (specific learned knowledge)</td>
</tr>
</tbody>
</table>

**Table 4-6: Differences between primitive and schema-based scene analysis**

Voluntary attention employs schemas such that anything being “listened for” is part of a schema. Where there is attention, there is schema-based segregation. Attention involves an experience of effort. The number of things that can be attended to is limited.

Separating effects due to primitive scene analysis versus those due to schema-based segregation in experiments is difficult since most experiments involve directing the participant’s attention.

Bregman makes a critical distinction between primitive processes that “partition the sensory evidence whereas schema-based attentional ones select from the evidence without partitioning it.” (Bregman, 1999: 669)

Primitive partitioning is symmetrical in that if low tones are separated from high tones, we can listen to either the low ones or high ones. Schema-based selection is not symmetrical. For example, hearing one’s name in a mixture of sounds does not make it easier to tell what the remainder of the mixture consists of.

The schema-based process looks at relations over a longer time span than the primitive one can.
4.5.3 The existence of a primitive process

To establish the existence of a primitive segregation or partitioning process, one could look for the following in the process:

- It must exist independently of learning.
- It must not care about the regularity of the sequence.
- It must operate independently of attention.

Conversely, the presence of any of these factors will indicate a schema-based approach.

4.5.4 Trajectories

Is a regular sequence of sounds (where the sequence is predictable) easier to recognize than an irregular one?

As we listen repeatedly to an auditory pattern, we learn the regularities in it. We can then anticipate segments of the pattern before they occur and integrate the sequence into a coherent mental representation. This preparation of our attention through repeated listening should assist the formation of sequential streams.

Experiments on listeners’ memories for regular and irregular sequences of tones have confirmed this theory. Since these tests involve memory, this kind of streaming is a schema-based process.

Other experiments have shown that “the regularity of a sequence does not effect the formation of streams” where listeners judged the number of streams without having to remember the patterns (Bregman, 1999: 670). Here streaming is formed by primitive processes.

Using a rising or falling series of pitches as an example of a predictable sequence, Bregman poses the question: “Can primitive scene analysis follow and segregate such trajectories from their acoustic contexts?” (Bregman, 1999: 670)
There is evidence both for and against such abilities in the formation of streams via trajectories.

**Evidence against stream formation via trajectories**

**Case 1**: A sound composite is made up of two simultaneous series of tones, one descending, one ascending (See Bregman, 1999: Fig. 4.2, Ch 4). Listeners find it hard to follow one sequence through the cross-over point. We tend to follow the descending series to the cross-over point, then we switch and follow the ascending one. So some preference of frequency range is taking precedence over linking by trajectory.

**Case 2**: A rising and falling glide tone has noise bursts inserted at peaks in the frequency pattern (See Bregman, 1999: Fig. 1.15, Ch 1). The auditory system restores the “missing” peak but not at the missing peak frequency. The peak frequency perceived is the highest frequency actually present, that is the highest one not deleted. The primitive restoration mechanism does not use the trajectory of the pitch contour to predict the missing points.

**Evidence for stream formation via trajectories**

**Case 1**: Segments of tone glides interrupted by a noise burst will be perceived as a single glide when the segments before and after the noise burst line up on a common trajectory.

**Case 2**: The task of reporting the order in a sequence is easier when tones follow a rising or falling trajectory. This may be due to the ease of encoding of the trajectory in memory rather than perceptual grouping. Effects of regularity are apparent in memory tasks.

Bregman concludes that “the primitive Gestalt-like grouping process does not make use of the fact a sequence may be rule-governed or may point to some future position of a sound in frequency or in time.” (Bregman, 1999: 672) In other words, rule-governed sequences are schema-based rather than Gestalt-based.
Trajectories can be both rule-based and Gestalt-based at the same time. The nearest neighbours in frequency and time within a trajectory sequence can follow a rule and maintain perceptual integrity simultaneously.

**Case 3:** Some theorists, who contend that streams are formed by attention, have argued that rhythmic regularity is responsible for the formation of streams since they involve predictability and predictability influences attention. Experiments with temporally irregular sequences involving high and low frequencies have produced streaming. Undoubtedly temporal regularity can lead to the recognition of patterns through schema-based segregation/integration. Are these forms of rhythmic trajectories?

Do we need a separate category for temporal patterning rather than applying the term trajectory?

What we can say is that the concept of trajectory is important, but its application is far from straightforward. We may also need a separate category we could call temporal patterning.

4.6 Primitive Auditory Organization in Music

Schema-based organization in the perception of music is an area where not a lot of research has been carried out, especially in relation to acousmatic music. Instead, attention here will be focused on Bregman’s work on primitive auditory organization in (mostly) tonal music. Since Bregman contends primitive auditory organisation is largely innate and built upon Gestalt principles, then what is relevant for tonal instrumental music should also be relevant for acousmatic music.

4.6.1 Primitive organization in music

Bregman is careful to point out that: “Both primitive scene analysis and complicated schemas play a role in our perception of music.” (Bregman, 1999: 674) However he restricts his discussion to the primitive processes only within his text.

Both the horizontal (melody) and vertical (harmony) dimensions in music require primitive organization – sequential organization and simultaneous organization.
In this section of his book, Bregman observes that scientific discovery of explicit factors influencing spectral integration could provide a rational basis for a theory of orchestration: “If based on principles of auditory organization, such a theory could be independent of particular musical styles.” (Bregman, 1999: 679) What is interesting for our study is the phrase “independent of musical styles”.

By analogy we could argue that the role of primitive scene analysis (as opposed to schema-based scene analysis) for tonal music should be largely transferable to acousmatic music since primitive organization involves innate heuristics and the application of Gestalt-style principles in a pre-attentive manner. According to Bregman’s influence, such heuristics and principles should apply in both domains (tonal music and acousmatic music).

Thus with suitable word substitutions (e.g. frequency for pitch, sound event for note, etc.), Bregman’s description of primitive organization in tonal music should translate into primitive organization in acousmatic music.

Let us test this proposition to create some principles for primitive scene organization in acousmatic music.

We must be careful to make the distinction between the experience of sound in the “real world” and the experience of music:

Natural hearing tries to avoid chimeric percepts, but music often tries to create them. (Bregman, 1999: 459-460)

The perception of music is different from the perception of the natural environment. Environmental sounds are segregated into streams that are created by specific sources, with mixtures being largely accidental or coincidental. In music, fusion and segregation must be carefully controlled to create new timbres from blended instruments, or to create a single melody across a number of instruments.

(Listeners) … must hear fictional sources of sounds that have qualities that emerge from the set of sounds being grouped. (Bregman, 1999: 674)

The grouping is not a real source, but what Stephen McAdams calls a “virtual source”. The virtual source in music has the same function as a real source in natural
environments: “Real sources tell a true story; virtual sources are fictional.” (Bregman, 1999: 460)

A “first principle” can be extracted from Bregman’s auditory scene analysis applied to music:

**Principle #1:** In music, listeners hear fictional sources with qualities emerging from a set of sounds being grouped.

Separate perceived events must relate to the musical architecture as a whole. Segregation occurs between different parts of a piece of music. Such segregation must be both strong and weak. It must be strong enough to enable perception of each melodic line and timbre, and weak enough to be able to establish the relations with other parts. A hierarchical form of organization is a good solution to creating a balance between such strong and weak forces. Bregman uses the structuring of a human hand as an analogy to the hierarchical structuring of perception. He notes that: “… we do not know whether these hierarchical structures are formed by primitive or schema-based approaches. I rather suspect it is the latter.” (Bregman, 1999: 675)

So although Bregman side-steps hierarchical organization and the higher-order forms of organization in music (scales, chord relations, etc.), hierarchical organization can be hinted at through the statement of a second principle:

**Principle #2:** Separate perceived events must relate to the musical architecture as a whole.

**4.6.2 Melody**

Melody is governed by sequential organization. A sequence with large frequency transitions in a short space of time will not remain perceptually coherent:

Small transitions in the fundamental frequencies (or pitches) of notes are much more common than large ones in traditional Western music, and when the speed has to be made very fast (as in trills, grace notes, and other ornamentation) the pitch jumps are made very small. (Bregman, 1999: 675)

Alternations of high and low tones will cause segregation between the two different registers and the perception of separate melodic lines (or streams) will result. This
phenomenon has been used as a compositional device, especially in the Baroque period.

Melody perception has been tested in the presence of distracter tones:

If the distracters fall within the same pitch range as the melody, the recognition is very hard. The further the distracters are in pitch from the melody, the easier recognition becomes. (Bregman, 1999: 676)

Transformations in pitch define a melodic form: “For example, a rising sequence of tones is perceived as a rising gesture.” (Bregman, 1999: 676) The tones defining the beginning and end of the rising melodic gesture must be perceived as being part of the same stream, i.e. transitions must be packaged with streams.

This makes sense in nature where transitions signal changes in sound sources - footsteps getting louder signal the walker is getting nearer. Here the notion of a single sound source is coupled with a single sound stream in our mind. Different sound sources will seem to have unrelated intensities. Meaningful transformations in the real world are identified with a single source and are represented as a single (sequential) stream. Bregman states:

We have presumably developed a perceptual system that looks for meaningful transitions within streams. Musical transformations have to be content to live within such a system. (Bregman, 1999: 676)

Changes in timbre can effect the integration of a melodic sequence, for example:

- Repeated and/or rapid changes in timbre can fragment a melodic sequence.

- Less rapid shifts in timbre can be used to delineate larger melodic units or “phrases”.

A distinction can be made between form-defining melodic notes and subordinate ornamental notes. Ornamental notes group with the form-bearing note they are subordinate to: “The Gestalt psychologists called this relation phenomenal dependency.” (Bregman, 1999: 677)
Gestalt-like principles of primitive scene analysis seem to constrain these emergent events:

For example, the dependent note must be very close to the anchor note in frequency and time. Otherwise the two will not group to form a larger event. (Bregman, 1999: 677)

The choice of which note is the anchor and which is the dependent one depends on factors such as duration, intensity, and rhythm. Dependent tones tend to “resolve” to stable anchor tones.

Bregman notes that: “When component sounds are grouped sequentially to form a larger sonic event their own individual properties are to some extent lost.” (Bregman, 1999: 677) This effect is less on the highest and lowest notes in a sequence than on those tones in the middle.

Notice in the above discussion on melody there is minimal mention of pitch, and in fact much of the discussion would be just as cogent if the word “sound” replaced the word “tone”. This will have implications when we discuss the organization of timbre. Let’s see what those implications are.

4.6.3 Timbre (Cause and Effect)

Bregman notes that timbre plays a role in the sequential organization of music. Timbre is a complex phenomenon and no one perspective can encapsulate its complexity. Bregman tackles timbre from several points of view, under separate headings.

Timbre as a cause of segregation and integration

As noted earlier, large and rapid changes in timbre can cause the formation of separate, parallel streams as happens with a “compound melody”.

Conversely, timbre can be used as sequential glue for musical phrases, and to delineate separate units or sections. Such delineation exploits scene analysis principles from nature where a sudden change in timbre usually implies a new event has begun. In contrast, a continuous change implies that a single event is changing (incrementally) in some way – through some form of transformation.
The dimensional approach to timbre

Bregman points out that there is a use of timbre that is stronger than the accentuation and reinforcement of melodic forms:

If we knew what the dimensions (of timbre) were, we could use movements along them to define transformations that could serve as structural elements in a piece of music. (Bregman, 1999: 678)

Bregman highlights at least four approaches to the study of the dimensionality of timbre:

1. Similarity studies – Similarity ratings between pairs of sounds has been used to try to determine the number and nature of dimensions. Some studies have produced three dimensions:
   
a. The brightness of the spectrum.

   b. The bite of the attack.

   c. The simplicity of behaviour of the harmonics over time.

Lerdhal (1987) has attempted some tests of the organization of music according to similar types of dimensionality. It is important to note that these results are for instrumental timbres and we have already seen that the dimensionality is much more complex for environmental sounds.

2. Acoustic dimensions – Differences in spoken vowels correspond to differences in formant frequencies (prominent peaks) in their spectra, especially the two lowest formants. Slawson (1985) has attempted to organize music according to these timbral dimensions.

3. Physical modelling – This approach assumes our brains are built to form descriptions of environmental events rather than sounds in the abstract. In this view, changes in sound attributes correspond to changes in dimensions or components within a mental model of a physical system that produces the sound.
Timbre as the result of fusion

Returning to discuss the relationship between timbre and scene analysis, Bregman reminds us that a certain paradox is represented where timbre influences scene analysis at the same time that scene analysis creates timbre. The identity of component sounds in a “scene” is the result of partitioning through scene analysis which allows us to hear different sounds at the same time. The use of timbre in music depends on an understanding of auditory fusion and segregation.

Knowledge of fusion and segregation principles can be used to assist the organization of musical texture, for example:

- A number of musical sounds can be fused to create a global timbre.
- A polyphonic texture can be created where two or more distinct melodic lines can be heard.

Music uses scene analysis principles in a number of ways. Here are some examples:

1. Separation of soloist from accompaniment:
   a. The solo part uses different pitches to the accompaniment at the same time (segregation based on harmonic series).
   b. When the same pitches are played, mistune the soloist slightly, e.g. using vibrato.
   c. The solo part occupies a different register to the accompaniment, e.g. “singing formant”.
   d. Minimize synchronization of onsets and offsets, e.g. use rubato.
   e. Use spatial location for separation.

2. The role of primitive scene analysis in counterpoint:
   a. Although schema-based principles are used, Bregman discusses primitive scene analysis only.
b. In polyphonic music, the parts must not be totally segregated or totally integrated.

c. Segregation between parts is improved by strong sequential organization within each part. Strong sequential organization is achieved through the following principles:

i. Small pitch changes favour sequential integration.

ii. For lines to remain distinct, parts should be well separated in pitch and should not cross.

iii. Segregation of parts is improved by weakening fusion of notes from different parts. Avoid “common fate” by prohibiting synchronous onsets and offset (different rhythms). Avoid parallel changes in pitch between two parts, i.e. encourage contrary motion or oblique motion. Avoid harmonic relations between simultaneous notes in different parts, e.g. whole number ratios like 2:1, 3:2, etc.

d. Vertical integration is achieved by violating the above rules.

Traditional definitions of timbre have characterized timbre as being determined by the qualities that distinguish one instrument from another. For Bregman, instruments have been replaced by sonic objects which are manifestations of auditory streams:

Timbre has come to mean the qualities that distinguish one sonic object from another, where sonic objects can no longer be identified with single instruments and correspond more to what I have spoken of earlier as auditory streams. (Bregman, 1999: 490)

4.6.4 Dissonance

Bregman distinguishes two types of dissonance: psychoacoustic dissonance and musical dissonance:

Psychoacoustic dissonance is the sense of roughness or unevenness that occurs when certain combinations of simultaneous tones are played. This sort of dissonance is not defined by the musical style. (Bregman, 1999: 681)
Roughness is caused when partials combine to produce a large number of beats at an unrelated rate. Simple ratios like 3:2 (ca. seven semitones) will sound smooth, complex ratios like 45:32 (ca. six semitones) will seem rough.

Musical consonance is defined by Bregman as a cognitive experience. Stable and unstable combinations of sounds are defined by a musical style. Unstable combinations are points of tension, whereas stable ones are points of rest: “By the terms musical consonance and dissonance, I am referring to this stability and instability.” (Bregman, 1999: 682)

Bregman notes that unstable combinations of tones often happen to be psychoacoustically dissonant in Western music. Vertical (psychoacoustic) dissonance between a combination of two tones can be reduced by capturing each into a different perceptual stream. Gestalt psychologists point out that perceived qualities belong to organized perceptual units.

Composers control such dissonant tones by:

1. Avoiding simultaneous start and stop times.
2. Capturing tones into separate streams by preceding each with tones close in pitch.
3. Capturing each into smooth and different trajectories.
4. Capturing each into its own repetitive sequence.
5. Using “polytriads” where notes from different triads are prevented from grouping with each other by establishing internal relationships with other notes in their own triad.

An increase in dissonance can be created by violating the above principles

4.6.5 Rhythm

Bregman does not write extensively on rhythm, but he does note that any sort of perceptual feature that depends on isolating some of the patterns in music can be influenced by primitive grouping.
On rhythm, Bregman contends: “Rhythms are relationships between auditory events in the same stream.” (Bregman, 1999: 683) Without the presentation of further evidence, it is hard to evaluate veracity of this statement.

Bregman does highlight some peripheral issues, for example a polyrhythm, played on a single instrument, can be separated into its two component rhythms by assigning widely separate pitches to each rhythmic element. But his discussion of rhythm is rather too brief.

4.6.6 Primitive Auditory Organization Postlude

The following quote encapsulates much of Bregman’s ideas on primitive auditory organization in music and provides us with a neat, short summary of this section:

… sequential and simultaneous organizations actually create certain aspects of musical experience. Sequential grouping creates rhythms and different aspects of melodic form, vertical grouping gives us not only the experience of chords but also the emergent qualities of simultaneous sounds, e.g. timbre, consonance and dissonance. (Bregman, 1999: 459)

4.7 Narmour and the implication – realisation model

In an article entitled The “genetic code” of melody: Cognitive structures generated by the implication-realization model, Narmour (1989) provides some insight into how expectancies may be created and interrupted by applying an implication-realization model to the field of melodic perception in tonal music. Narmour claims “that the analysis of melody rests on the perception of implications of continuation and reversal. Continuation is said to be governed hypothetically by the bottom-up Gestalt laws of similarity, proximity, and common direction (common fate); whereas reversal is hypothesized as a symmetrical construct.” (Narmour, 1989:45)

Narmour begins by noting that perceptual groupings are created in chunks defined by beginning and ending notes that have some degree of stability. The principal causes of melodic grouping are duration, harmony and meter. They create some degree of closure and Narmour characterizes such durational, harmonic, or metrical closure as an interference on melody: “… what is interrupted in a melodic line is the realization
of implication.” (Narmour, 1989:46). Here Narmour uses the term implication in the way that psychologists use the term expectancy.

### 4.7.1 Implication of continuation

How are such implications (or expectancies) created in the case of melody? Narmour notes:

> The implication-realization model hypothesizes that intervallic continuation, registral direction, and specific pitch (when mode is known) are all separately subject to cognitive prediction and thus dependent on the laws of implication and expectancy. (Narmour, 1989:46)

What are the psychological foundations of the influences of interval, register and pitch? Narmour claims they are governed from the bottom up by the Gestalt laws of similarity, proximity, and common direction (actually common fate). For example:

> … the pitches F-G imply the closely proximate pitch of A (assuming mode is known); the major second of F-G implies another similar interval (in this mode, another major second); and the ascending F-G leads to directional expectations of a continuing upward register. (Narmour, 1989:47)

The bottom-up Gestalt laws have certain advantages over the top-down Gestalt laws of good continuation, good figure and best organization in that they are measurable and open to empirical testing. Narmour also notes that the bottom-up Gestalt laws are resistant to learning and may be innate.

### 4.7.2 Definitions

Narmour invents his own shorthand notation to describe the concepts he introduces to unravel how such a melodic processing may be achieved in a practical sense. So pitch proximity is described as two tones the distance of a major second apart or smaller, intervallic similarity as adjacent intervals differing by a minor third or less, and common direction as up followed by up, for instance. Narmour hypothesizes that all intervals of a perfect fourth or less imply a continuation of intervallic similarity and registral direction. Such bottom-up constants exist in the absence of top-down stylistic interference.
Patterns of intervallic similarity and registral continuation are described as registral-intervallic processes, or just processes, and are identified by a capital letter P. Iteration, or duplication, is symbolized by the letter D. Narmour also introduces the notions of partially realized patterns of intervals (IP) and partially realized, zigzagging, registral intervals of intervallic duplications (ID). Additionally, implications of pitch and interval can be denied but implication of registral direction can be realized (VP). In other words we can experience partial surprises.

### 4.7.3 Implication of reversal

Some patterns with melody are capable of creating closure by themselves without the assistance of durational cumulation, resolution of dissonance, or metric differentiation. One type is intervallic registral reversal. Narmour hypothesizes that when a listener hears a large interval (>P5) “he expects, all other things being equal, a change in registral direction (A+B) and a differentiation of interval”.

(Narmour, 1989:49)

Narmour supports the implication of continuation with certain bottom-up Gestalt laws, but he can provide no psychological evidence to support his implication of reversal hypothesis. Instead, he elects to treat melodic reversal as a “symmetrical construct” within the implication-realization model.

### 4.7.4 Narmour’s parametric scale

The eight basic melodic structures that Narmour identifies are: process (P), intervallic process (IP), registral process (VP), duplication (D), intervallic duplication (ID), reversal (R), intervallic reversal (IR), and registral reversal (VR).

Table 4-7 formalizes, on a parametric scale, the bottom-up hypotheses of how small intervals imply process and large intervals imply reversal.
Table 4-7: Narmour's parametric scale

<table>
<thead>
<tr>
<th>u</th>
<th>m2</th>
<th>M2</th>
<th>m3</th>
<th>M3</th>
<th>P4</th>
<th>A4/d5</th>
<th>P5</th>
<th>m6</th>
<th>M6</th>
<th>m7</th>
<th>M7</th>
<th>(P8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sameness similarity</td>
<td>differentiation ………</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a+a)</td>
<td>(a+a’)</td>
<td>(a+b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>continuation implied</td>
<td>(threshold)</td>
<td>reversal implied</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>weak implication</td>
<td>←---------------→</td>
<td>strong implication</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Narmour claims that this parametric scale is a “perceptual hypothesis in the implication-realization model against which degree of implication (in P, IP, ID, and VP), degree of closure (in R, IR, and VR), and degree of surprise are measured and evaluated.” (Narmour, 1989:51)

4.7.5 Retrospective realizations

Implication and realization involve prospective perceptual evaluation, but Narmour notes that listeners retrospectively evaluate realizations by looking back over what they have experienced: “If this were not so, listeners could never correct their perceptual ‘mistakes’, ‘re-write’ their short-term memories, or learn new implicative associations.” (Narmour, 1989:52)

Thus Narmour adds eight retrospective categories to his original eight prospective structures giving him a palette of sixteen basic structures.

4.7.6 Repertoire examples

Narmour applies his implication-realization model to a number of short melodic examples from the repertoire: Mozart, Debussy, Schoenberg, and Bach. He uses his shorthand notation to illustrate examples of the basic structures.

Narmour then introduces examples of combinations: “where the end of one pattern shares at least two tones with the beginning of another” (Narmour, 1989:54) and chains where “it is possible for three or more of the sixteen basic prospective and
retrospective types of structures to join into longer structural chains.” (Narmour, 1989:55)

4.7.7 Encoding

With respect to encoding, Narmour makes the following claims:

- The encoding is completely independent of style-structural schemata.

- A fairly wide-reaching musical expertise could be based on a very few principles. “Indeed, it may be the case that, of all the arts, syntactical structure in music is the least dependent on learning.” (Narmour, 1989:58)

- A naïve listener would have little trouble recognizing low-level structuring and high level structural tones of a melody.

- Cognitive assimilation takes place through horizontal processing and vertical construction of hierarchical levels.

4.7.8 Narmour’s Conclusions

Narmour identifies some areas for further research:

- How does low-level implication influence high-level implication?

- Are definitions of intervallic similarity and difference constant from listener to listener and style to style, or do they vary according to degree of learning?

- Does parametric scaling in music bear any direct connection to specific neuropsychological structures?

In Narmour’s work, we have seen an attempt at an explanation of how perceptual processing may influence the structures that are present in music, and then how those structures may imply certain continuations or realisations in listeners minds. Narmour has provided us with a very real demonstration of how a syntax can arise from our perception of events and how that syntax can influence our perception of subsequent events.
We can also get a picture that many researchers have not found it easy to unravel the relative contribution of top-down and bottom-up processes. Narmour accentuates his bottom-up approach, but his inclusion of patterning implies some top-down influences too.

4.8 Summary

How can the examination of signs, symbols, semantics and syntax help with an analytical approach to acousmatic music?

This chapter has attempted to concentrate more on schema-based interpretation of music. The work of Dowling and Harwood (1986) on emotion and meaning has provided a useful outline on the interpretation of signs and symbols in Western tonal music. They are careful to distinguish between the representation of emotions and the direct elicitation of emotions. Music can do both, although not necessarily at the same time, and the emotion represented is not necessarily the emotion induced.

Dowling and Harwood suggest that musical meaning depends on symbolic representation and musical symbols function through the application of a musical syntax. They use a classification system for signs developed by Pierce (1931-1935) that has three basic types: index, icon, and symbol.

Indexical representation refers to the direct association of a musical event with a non-musical event. Iconic representations use patterns in the music to mimic the ebb and flow of emotional tensions and relaxations. Context is important to the interpretation of iconic emotional meaning in music. Context can be learned through familiarity acquired through repeated listening during rehearsal or repeated exposure to recordings.

The meaning of indexical or iconic signs can often be read from the sign directly, but the meaning of a musical symbol arises from its place in the syntax of a piece and the way it relates to a musical style. A symbol, therefore, is a sign whose significance depends on its relationship with other symbols.

How do patterns of symbols actually elicit emotions or convey meaning? Dowling and Harwood use Mandler’s theory of emotion which links arousal with meaning. When
listening to music, interruption to ongoing schemata causes arousal. The arousal triggers cognitive interpretation, or the search for meaning. Together the arousal and the interpretation may produce an emotional experience. According to this theory, Dowling and Harwood assert that listeners’ schemata accommodate to a style, and that composers need to extend the stylistic boundaries in order to create the new interruptions that are required to cause arousal.

To speculate on how these mechanisms may operate in acousmatic music, we first concentrate on the recognition of signs and symbols in acousmatic music, and then explore the relationship between those signs.

To help with the recognition of sound events in acousmatic music, Gygi’s research on the identification of environmental sounds has been drawn upon, since the recording and manipulation of environmental sounds is endemic to this genre of music. Gygi’s research reveals a lot about the factors involved in the identification of environmental sounds, in particular the acoustic factors that may be responsible for recognition. Gygi’s results show that: the important frequencies for identification are in the range 1200-2400 Hz; environmental sounds can be severely filtered and still be identifiable; environmental sounds have many acoustic cues due to their complex spectral-temporal composition, and this redundancy assists robustness to signal degradation; both central and peripheral factors are involved in the identification; environmental sounds are more “speech-like” than laboratory-generated sounds; future studies should not only look at acoustic determinants, but also consider the interaction of cognitive and sensory factors.

Gygi also presents results of similarity studies conducted at the Hearing and Communication Laboratory at Indiana University. He points out that for “meaningful” sounds, there needs to be a consideration of their “psychological space”. One experiment attempted to uncover the structure of such a psychological space by using multidimensional scaling (MDS) procedures in order to discover what dimensions describe such a space. A two-dimensional MDS solution was derived, and certain clusterings were found: animal sounds, more rhythmic sounds (clapping, ping pong ball bouncing), and water-based sounds tend to cluster. Gygi searched for a meaningful interpretation of the two dimensions, and measured over twenty acoustic
variables. Based on the clusterings, it was hypothesized that some measures of ‘pitchiness’ or harmonicity might account for the ordering on the first dimension. Dimension 2 is not quite so straightforward, but it was thought that some measure of rhythmicity might account for the variance on Dimension 2. The highest correlations for Dimension 1 were: Spectral spread (Standard deviation of the spectrum, SD of the spectral centroid velocity); and confidence in the pitch of the signal (Mean pitch salience).

Correlations with the second dimension were not so clear cut. There were correlations with some variables, but the correlations were low. These low correlations were with two main types of variables: Measures of rhythmicity and continuity (Number of peaks, Pause corrected RMS/longterm RMS - reflecting the amount of silence, Autocorrelation - reflecting some periodicity); and measures indicating high frequency energy (Spectral mean, Energy in high frequency bands, Median pitch).

The contribution of syntactic and semantic factors to the recognition of sound events is the subject of the work of Howard and Ballas. Working firstly with patterns of sound in the form of pure tones and then with environmental sounds. They constructed grammatical sequences using a somewhat arbitrary finite state grammar.

In their first experiment with sequences of pure tones, Howard and Ballas found that the “grammatical” group performed significantly better in the recognition of target patterns compared to the “non-grammatical” group. Their most startling finding was that listeners in the grammatical group had actually learned something about the syntactic rules used to generate the target patterns during the administration of the test alone!

Their second experiment used sequences of complex environmental sounds and the performance of the grammatical group was considerably lower than in the first experiment. The results seem to show that some form of semantic processing was interfering with the classification task.

So they designed a third experiment that explored the four combinations of two variables: a syntactic variable and a semantic variable. Where the variables coincided, that is where the syntactic sequence made semantic sense, the grammatical/semantic
group performed significantly better than the non-grammatical/semantic group and both the non-semantic groups.

To test the relative contribution provided by acoustic, semantic, syntactic and other factors to the identification of common everyday sounds, James Ballas designed a series of experiments that measured identification times. The first experiment showed that acoustic factors can account for 50% of the variance in identification time. The effect of familiarity of sounds to the participants was examined in experiment two. Ballas showed that a certain combination of acoustic factors and ecological frequency accounted for 75% of the variation in identification time. The most significant model comprised of: harmonics in continuous sounds and similar spectral bursts in non-continuous sounds, the average frequency of the spectrum, the ratio of burst duration to total duration, and ecological frequency.

A third experiment examined the contribution that perceptual and cognitive judgements make to sound identification. Ballas found that sound identifiability was related to: The ease with which a mental picture of the sound is formed; context independence; the familiarity of the sound; similarity with mental stereotype; ease in using verbal description of the sound; clarity of the sound.

The issue of ‘categorisation’ was addressed during analysis with Ballas performing a cluster analysis which resulted in the creation of a tree diagram depiction. Initially, Ballas identified four categories: water sounds, signalling and danger sounds, door sounds and sounds of modulated noise, sounds with transient components. He also noted that “signal sounds” (telephone, door bell, bugle) were more easily recognised than their ecological frequency would suggest, and a hybrid theory should incorporate the effect of contextual information.

Having dissected the process of sound event recognition and examined the most salient factors in everyday sound identification, we then moved on to consider the relationship between sound events in acousmatic music.

We initially looked at the issue of the simultaneous and sequential grouping of frequencies and sounds. To assist the discussion of this issue, we returned to
Bregman’s theories of auditory stream segregation and the differences between primitive segregation and schema-based segregation.

Bregman writes extensively on primitive auditory organisation in music and makes these two important observations concerning qualitative differences between listening to an ordinary, everyday sound environment and listening to music:

1. In music, listeners hear fictional sources (virtual sources) with qualities emerging from a set of sounds being grouped.

2. Separate perceived events must relate to the musical architecture as a whole.

The second principle hints at the hierarchical nature of organisational structures in music.

Bregman expands on the primitive organisation principles that may form melodic elements. For example, he makes the distinction between form-defining melodic notes (anchors) and subordinate ornamental notes. Ornamental notes group with the form-bearing note they are subordinate to (dependent tones “resolve” to stable anchor tones). Timbre is examined as a cause of segregation and integration, as a multi-dimensional property to be explored in its own right, and as an emergent property that is more like a global timbre created from a hybrid of sources. Here Bregman discusses musical concepts such as polyphony and how the concepts may emerge from sequential organisational phenomena.

Dissonance, an important concept in music, is classified into two types: psychoacoustic dissonance and musical dissonance. Bregman then provides recourse to Gestalt principles to show how composers controlled dissonance. He does not write extensively on rhythm.

Having explored recognition and grouping as they may be manifest in musical context, we then proceeded to find research that examines the connection between arousal and meaning in music. Recall that the Dowling and Harwood model of emotion links the interruption of ongoing schemata to arousal, which then triggers cognitive interpretation or the search for meaning. Narmour (1989) provides some insight into how expectancies may be created and interrupted by applying an
implication-realization model to the field of melodic perception in tonal music. Narmour claims “that the analysis of melody rests on the perception of implications of continuation and reversal. Continuation is said to be governed hypothetically by the bottom-up Gestalt laws of similarity, proximity, and common direction (common fate); whereas reversal is hypothesized as a symmetrical construct.” (Narmour, 1989:45)

Narmour notes that perceptual groupings are created in chunks defined by beginning and ending notes that have some degree of stability. The principal causes of melodic grouping are duration, harmony and meter. They create some degree of closure and Narmour characterizes such durational, harmonic, or metrical closure as an interference on melody: “… what is interrupted in a melodic line is the realization of implication.” (Narmour, 1989:46) Narmour’s recourse to the Gestalt laws of similarity, proximity, and common fate may prove useful when we discuss implications in acousmatic music.

Having surveyed such an extensive landscape of research concerning the perception of everyday sounds and the perception of some relevant phenomena in tonal music, where there has been some research, in the next chapter we will see how this can all come together in the assembly of a uniform framework for the analysis of acousmatic music which has some cognitive psychological basis.
Chapter 5 : Creating the Analytical Framework

5.1 Introduction

Chapter 3 began with the “bottom-up” approach of examining each sound event in turn and Bigand’s notion of Event Structure Processing was found to be useful. It was also observed that “top-down” considerations had to be included within the application of his model. Chapter 4 started with a “top-down” approach to Signs and Symbols, Semantics and Syntax but it was quickly found that “bottom-up” considerations had to be included when discussing segmentation and integration within the auditory system.

Bigand (1993) presented a model of event structure processing whereby a “sliding window” of listener perception analyses the work on a moment by moment basis, and a matching process compares salient features with musical knowledge representations in long-term memory. It implies that where relevant knowledge representations are recognized, or activated, this activation is somehow retained for processing at some time in the future. For a listener processing musical auditory information, Bigand’s model of event structure processing identifies six stages: reading the acoustic surface; auditory image formation; time-span segregation; time-span reduction; prolongation reduction; and symbolization processes. In this chapter, the aim is to create a framework that is a synthesis of the Event Structure Processing model with the Signs and Symbols, Semantics and Syntax ideas derived from the work of Bregman (1999), Gygi (2001), Howard and Ballas (1980), Ballas (1993), (Narmour, 1989), Dowling and Harwood (1986), Smalley (1994), and Pierce (1931-35).

The study’s basic assumptions need to be re-iterated here too:

1. That a hierarchical means of organization and mental representation operates. If long-term structures and relationships are to be established and perceived in acousmatic music, then there must be some form of hierarchical organization operating. This need for a hierarchy stems from a need for information reduction resulting from the limits of the psychological present (short-term memory).
2. That there is a perceptible scheme of alternating periods of tension and relaxation, stability and instability, or “discourse” within the works studied.

3. That metric organization is not as significant for the type of music we are going to consider in the course of this specific investigation. In other words we will select works in which a discernable pulse is not the predominant factor in the unfolding of the work. We are seeking to hold one variable constant in order to consider whether hierarchical structuring is possible when sound events (or timbral units) are the primary structuring force. Structuring in time will still be important, but we will initially consider works where no strong pulse is present.

5.2 Defining the SIAM Framework

The following procedure for the analysis of acousmatic music was derived from the synthesis of top-down (knowledge driven) and bottom-up (data-driven) views since it has been found impossible to divorce one approach from the other. The procedure is also a synthesis of the primitive auditory scene analysis approach of Bregman (1999), with the research on acoustic, semantic, and syntactic factors in the perception of everyday environmental sounds carried out by Gygi (2001), Howard and Ballas (1980), and Ballas (1993), together with theories of implication-realization (Narmour, 1989) and symbolic interpretation (Pierce, 1931-35).

This methodology has been given the acronym “SIAM”, which stands for Segregation, Integration, Assimilation and Meaning. The procedure can be summarized as consisting of a number of stages and steps, as shown in Table 5-1. The “Section Reference” column in the table refers to the section within this thesis where the relevant research is detailed.
### Segregation of sonic objects

1. Identify the sonic objects (sound events).  
   - McAdams (1993)  

2. Establish the factors responsible for identification (acoustic, semantic, syntactic, and ecological) and their relative weightings (if possible).  
   - Vanderveer (1979), Gygi (2001), Howard & Ballas (1980), Ballas (1993) (4.3.2-4.4.4)

### Integration - Horizontal

3. Identify streams (sequences, chains, or patterns) consisting of sonic objects linked together, which function as a unit that we could call a “pattern”. The role of “trajectories” should also be considered. (The term “gesture” has sometimes been used in this context.)

4. Determine the causal linkages between the sonic objects within the chain-type pattern.

5. Determine the relationships between “pattern objects” – if this level of syntax exists. This amounts to the investigation of higher-order relationships within a “hierarchy”.

6. Consider local organization in time – pulse, beat, accent, rhythm, and meter.

7. Consider the horizontal integration of pitch, including emergent properties relating to timbre (vertical overlap).

### Integration – Vertical

8. Consider vertical integration as a cause of timbre creation and variance:
   - a. Timbre as a cause of integration and/or segregation.  
     - Bregman (1999)  

---

The Development of a Cognitive Framework for the Analysis of Acousmatic Music 130
### Table 5-1: The SIAM Framework

<table>
<thead>
<tr>
<th>SIAM Stages</th>
<th>Research</th>
<th>Section Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>b. The dimensional approach to timbre, including emergent properties relating to pitch.</td>
<td>Lerdahl (1987), Slawson (1985)</td>
<td></td>
</tr>
<tr>
<td>9. Also consider vertical integration or segregation in terms of the potential for psychoacoustic dissonance and musical dissonance (and consonance).</td>
<td>Bregman (1999)</td>
<td></td>
</tr>
</tbody>
</table>

**Assimilation and Meaning**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>10. Consider the nature and type of discourse on the source-cause dominant (semantic) to typological-relational dominant (syntactic) continuum, and the way it varies over time.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Consider global organization in time – identify formal structures, like sectional or continuous organization, and the nature of the relationships between musical sections, i.e. higher order hierarchical relationships. (Time-span reduction and prolongation reduction)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Consider implication-realization, expectation-interruption, arousal and meaning throughout the work.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**5.2.1 Segregation of sonic objects**

The first two steps in the methodology are the identification of sonic objects (or sound events), and establishing the factors responsible for identification.

Building on the work of Bregman, McAdams (1993) has postulated several stages in the recognition of sound sources and events. First of all there are the first two stages of sensory transduction and auditory grouping of frequencies – both simultaneous and
sequential. The third stage involves the analysis of auditory properties and features. McAdams focuses on the micro-temporal and macro-temporal properties of sound events to suggest that each of these categories reveals its own set of features which carry certain implications.

Auditory property analysis provides a group of abstract properties which act as input to the matching process with representations in memory. Categorical matching or lexical activation give rise to associated knowledge of properties of the sound “class” including its relationship to the listener and the local environment.

Categorical matching and activation of associations establishes a “sense” of the sound. After activation of a representation or category in memory has taken place, further activation of the listener’s lexicon of names, concepts and meanings associated with the “class” of sound events may occur: “Once language is available, however, the recognition process also gives access to a verbal lexicon that allows the listener to name or describe the event verbally.” (McAdams, 1993: 154-155)

There may be many factors responsible for categorical matching so the next step in our methodology seeks to explore those factors and the possible contribution made by each factor in sound event recognition.

5.2.2 Factors in the identification of environmental sounds

Acoustic factors

The work carried out by Vanderveer (1979) and Gygi (2001) was an extension of the work in the field of “ecological acoustics”. Their similarity studies provide a basis for establishing what acoustic factors might be involved in sound event recognition.

Gygi points out that for “meaningful” sounds, there needs to be a consideration of their “psychological space”. He attempted to uncover the structure of such a psychological space by using multidimensional scaling (MDS) procedures. A two-dimensional MDS solution was derived, and Gygi searched for a meaningful interpretation of the two dimensions. The highest correlations for Dimension 1 were spectral spread, and confidence in the pitch of the signal. Correlations with the second
dimension were not so clear cut. The most significant correlations were to do with measures of rhythmicity and spectral mean.

Gygi’s identification studies gives us a basis for isolating acoustic features of environmental sounds that we may want to single out for analysis.

**Syntactic and semantic factors**

The SIAM model draws upon the work by Howard & Ballas (1980) to shed light on any semantic (knowledge of source events) or syntactic (temporal structure) factors that may influence sound event recognition. What is important about their series of experiments is that they are concerned with patterns of sounds (i.e. sequences of sounds) – both pure tones and environmental sounds.

One of their most interesting findings was that the grammatical group performed significantly better than did the non-grammatical group, and that listeners in the grammatical group had actually learned something about the syntactic rules (the grammar) used to generate the target patterns - during the actual tests.

**Relative contribution of acoustic, semantic, syntactic and other factors**

So far we have examined some of the acoustic factors and some of the semantic and syntactic factors that may influence sound source identification. The SIAM framework also needs to take into account the relative contributions of these factors.

The experiment conducted by Ballas (1993) showed that the combined acoustic factors can account for up to 50% of the variance in identification time, and three quarters of the variance in identification time can be related to acoustic variables and ecological frequency.

Ballas found that sound identifiability is related to:

1. The ease with which a mental picture of the sound is formed.
2. Context independence.
3. The familiarity of the sound.
4. Similarity with mental stereotype.
5. Ease in using verbal description of the sound.

6. Clarity of the sound.

Sound clusters (categories) were formed on the basis of the factor dimensions:

1. Sound identifiability.
2. Timbre.
3. Oddity.

Now that segregation and the factors involved in the recognition of sonic objects has been considered, the details of horizontal and vertical integration can be attended to.

5.2.3 Integration - Horizontal

The next phase of the analytical procedure involves the identification of sequences (streams or chains). Such a sequence would consist of sonic objects linked together and it would function as a single unit that we could call a “pattern”. The role of “trajectories” should also be considered. The term “gesture” has sometimes been used in this context.

Next the linkages between the sonic objects within each chain-type pattern must be determined, including which of the relevant factors are operating, and the relative weightings of those factors.

Then the relationships between “pattern objects” must be determined – if this level of syntax exists. This amounts to the investigation of relationships within a “hierarchy”.

Local organization in time must also be considered – pulse, beat, accent, rhythm, and meter. This phase can be like a melodic analysis in tonal music. Transformations in pitch or frequency may define a melodic form (or gesture), or pitch can be treated as a quality that is emergent from the frequency spectrum of several sonic objects.

Bregman’s ideas on melodic organisation are useful when considering horizontal integration (see Section 4.6.2). Briefly recapping here - a sequence with large frequency transitions in a short space of time will not remain perceptually coherent.
Alternations of high and low frequencies will cause segregation between the two different registers and the perception of separate streams will result.

Changes in timbre can effect the integration of a horizontal sequence, for example:

- Repeated and/or rapid changes in timbre can fragment a sequence.
- Less rapid shifts in timbre can be used to delineate larger horizontal units or “phrases”.

A distinction can be made between form-defining sequential events and subordinate ornamental sound events. Ornamental events group with the form-bearing event they are subordinate to. Gestalt-like principles of primitive scene analysis seem to constrain these emergent events. The dependent event must be very close to the anchor event in frequency and time, otherwise the two will not group to form a larger event.

The choice of which event is the anchor and which is the dependent one depends on factors such as duration, intensity, and rhythm. Dependent events tend to “resolve” to stable anchor events.

**Pitch**

The horizontal integration of pitch can be considered, where appropriate. There is a certain overlap in viewing pitch in horizontal terms (like successive notes in a melody) and pitch-like characteristics that emerge through dynamic combinations of frequencies (vertical integration and segmentation). Smalley refers to this chimera as the pitch of timbre and the timbre of pitch (Smalley 1994:40).

**Trajectories**

When considering horizontal integration, the phenomenon of the trajectory needs to be considered. This phenomenon was discussed in Section 4.5.4 where it was found that trajectories can be rule-based, Gestalt-based, or both. For example, the nearest neighbours in frequency and time within a trajectory sequence can follow a rule and maintain perceptual integrity simultaneously.

The concept of trajectory is important, its application is far from straight forward, but it may be a strong feature in some acousmatic works.
5.2.4 Integration - Vertical

In this next stage, vertical integration as a cause of timbre creation and variance is considered, that is “spectral fusion”. Vertical integration or segregation in terms of the potential for psychoacoustic dissonance and musical dissonance must also be examined.

Timbre and texture

Timbre is a complex phenomenon and no one perspective can encapsulate its complexity. The role of timbre was discussed in detail in Section 4.6.3, so only the main concepts pertaining to the SIAM framework will be repeated here.

Timbre as a cause of segregation and integration

Timbre may be a cause of segregation. For example, large and rapid changes in timbre can cause the formation of separate, parallel streams as happens with a “compound melody”. Conversely, timbre can be used as sequential glue for musical phrases, and to delineate separate units or sections.

The dimensional approach to timbre

Bregman (1999) highlights several approaches to the study of the dimensionality of timbre, and some composers have explicitly used dimensionality in their works. Therefore the consideration of timbre dimensionality forms a part of the SIAM framework.

Timbre as the result of fusion, texture as a result of contrasting timbres

The identity of component sounds in a “scene” is the result of partitioning through scene analysis which allows us to hear different sounds at the same time. The use of timbre in music depends on an understanding of auditory fusion and segregation.

Knowledge of fusion and segregation principles can be used to assist the organization of musical texture. These principles can be adapted for the analysis of texture in acousmatic music (see Section 4.6.3).
Dissonance

Bregman distinguishes two types of dissonance: psychoacoustic dissonance and musical dissonance. Psychoacoustic dissonance is the sense of roughness caused when partials combine to produce a large number of beats at an unrelated rate. Musical consonance is defined by Bregman as a cognitive experience. Stable and unstable combinations of sounds are defined by a musical style. Some principles for the consideration of psychoacoustic and musical dissonance are defined in Section 4.6.4.

5.2.5 Assimilation and Meaning

In this final stage we consider global organization and formal structures. This is also the stage in which we discuss emotion, arousal and meaning, implication-realization, and the nature and type of discourses operating within the work.

In Section 4.5.2 the comparison was made between primitive versus schema-based segregation. Bregman (1999) states that to establish the existence of a primitive segregation or partitioning process, one could look for the following in the process:

- It must exist independently of learning.
- It must not care about the regularity of the sequence.
- It must operate independently of attention.

Conversely, the presence of any of these factors will indicate a schema-based approach. The schema-based process looks at relations over a longer time span than the primitive one can.

In Section 4.2, the concepts of signs, symbols, emotions and meaning in music were highlighted. In considering assimilation and meaning, all of these concepts need to be brought together in a comprehensive way.

Figure 5.1 represents many of the ideas gleaned from relevant research and seeks to cross-reference the approaches through the notion of “listener interpretations” and “listener attention”.
Figure 5.1: Listener interpretations of acousmatic music

The diagram presents four different styles of continua that could represent four styles of listener interpretation, and each continuum is represented as a dichotomy with its extremes noted at each end. Although they are shown as four separate layers, there is not necessarily vertical correspondence between the layers, and a listener may move freely around this interpretive space, as a single work is unfolding. The listener’s attitude and attention should be considered to be quite dynamic (as represented by the movable human figure in the diagram).

**Indexical to Symbolic Continuum**

A listener may describe his/her reaction to a work along the “semiotic” continuum where references extend from the indexical to the symbolic. References create simple associations, at one extreme, to directly eliciting emotions at the other extreme. A representation of the “form” of emotional ebb and flow lies in between (iconic).

**Source-cause Discourse to Typological-relational Discourse Continuum**

Listeners influenced by the “acousmatic music culture” may adopt the language provided by writers such as Smalley (1994), and seek to explore the continuum constructed according to the type of discourse that may be dominant at any given time.
Source-cause discourse is dominant when sounds from the natural world are identified according to their perceived source, and are related somehow by their everyday meaning. These sounds will have strong source-bonding, but as source-bonding decreases then acoustic factors begin to dominate and the work adopts a typological-relational discourse based upon more abstract relationships within a network governed by implications and realizations.

**Semantic Network to Syntactic Network Continuum**

Closely related is a continuum defined by a cognitive psychological vocabulary. At one end, recognised “individual sound events” will be related to each other via a network of semantic factors. As we move along the continuum towards the more abstract, physical recognition factors will still have ecological implications of the likely sources of “sound classes”. Spectral and temporal acoustic factors will tend to dominate as we move further to the right. At the right hand extreme, a syntactic network is formed from sounds which exhibit patterns determined by cultural experience.

**Concrete to Abstract Continuum**

The continuum at the bottom of the diagram differentiates between “Environmental Space” associated with the experience of “real-world” sounds, and “Musical Space” associated with the “fictional sound world” created by the musical work. Note that there is scope for overlap between these worlds so that the composer can play with ambiguities and chimerical properties resulting from the mixing of the two worlds. This continuum extends from the purely concrete reproduction of the environment, with the listener’s location within the artificial “mock” environment determined by environmental listening considerations, to the totally abstract world of a fictional sound space with fictional sources where the listener’s place within the “musical” space is determined by cultural and other considerations.

The above scheme provides a unified space for the discussion of concepts pertinent to acousmatic musical works, and it facilitates consideration of overall trends and features of such works, but the details must also be elucidated.
5.3 Conclusion

While the SIAM framework methodology has been derived from auditory scene analysis principles and research on environmental sound recognition, it is important to recall that the perception of the environment and the perception of music are fundamentally different. It was noted in the previous chapter that environmental sounds are segregated into streams that are created by specific sources, with mixtures being largely accidental or coincidental. In instrumental music, for example, fusion and segregation must be carefully controlled to create new timbres from blended instruments, or to create a single melody across a number of instruments. In acousmatic music, similar techniques are employed to blend sound events to create interesting hybrid timbres.

It was also noted in the previous chapter that segregation occurs between different parts of a piece of music. Such segregation must be both strong and weak. In traditional instrumental music, it must be strong enough to enable perception of each melodic line and timbre, and weak enough to be able to establish the relations with other parts. A hierarchical form of organization is a good solution to creating a balance between such strong and weak forces. We must consider the interplay of strong and weak forces in the chimerical world of acousmatic music too.

While the initial approach is one of deconstruction – tearing the musical materials apart, the acousmatic work must be put back together again in order to get a more holistic view. Time-span reduction and prolongation reduction will permit the consideration of higher-order hierarchies and the examination of different listener interpretation types.

What remains is to test this methodology thoroughly by applying it to an extended repertoire work – Denis Smalley’s Wind Chimes.
Chapter 6: Applying the Analytical Methodology

Previous chapters have described the development of a procedure for the analysis of acousmatic music which was derived from the synthesis of top-down (knowledge driven) and bottom-up (data-driven) views. This methodology has been called the SIAM framework for: Segregation, Integration, Assimilation and Meaning. The procedure has been summarized in Table 5-1.

This chapter reports on the application of the framework to the analysis of Denis Smalley’s Wind Chimes. (Smalley, 2004a).

6.1 Aims

The work Wind Chimes was chosen for analysis as it is a classic example of the genre described as acousmatic music, and while a recent article by Young (2004) considered some aspects of the work, the aim of the present study is to apply the Segregation, Integration, Assimilation and Meaning framework (SIAM) in an analysis of the work in order to test the appropriateness and usability of the SIAM framework, and secondly to discover insights into the work itself, such as the syntactic forces that may operate between sound events within the work. This process is somewhat like attempting to discover the “entities” and “relationships” present.

In order to work through this analytical framework, the questions of how to record the researcher’s observations and how to communicate any special insights both arise. So to record the results of the segmentation, integration, assimilation and meaning process, a Flash-based interactive was developed to provide a dynamic visual representation along with synchronized playback of the recorded work being analysed. The following sections describe previous methods of representation, the practical analytical methodology, and the design and development of the “interactive study score”.

6.2 Previous Methods of Representation

The “problem of the score” in the analysis of electroacousmatic music has been well documented (Hirst, 1995). There are traditionally two types of what may be called a
"score" in electroacoustic music - a list of operational data used to generate the piece, or a sketch of the musical effects obtained. The former is prescriptive, the latter descriptive. Descriptive scores have been criticised as being rather crude and approximate, but in order to communicate concepts discovered in the analysis of electroacoustic music, words and pictures are the main currency of communication. So is there a form of representation than can assist the communication of ideas and details pertaining to acousmatic musical works?

A number of writers have put pen to paper, and drawn pictures, in order to try and represent aspects of a fundamentally aural music tradition. Lewis (1998) created structural plans in a spreadsheet-like format, produced bar graphs of sound types, graphs of frequency of sound type changes and an elaborate graphical score of Dhomont’s Novars. In the same journal volume, Dack (1998: 111) created some beautiful frequency versus time transcriptions of Stockhausen’s Kontakte, Emmerson (1998: 156) graphed the left-right spatial disposition of instrumental motifs in Risset’s Songes, Roy (1998: 180-183) defined a whole lexicon of terms and symbols relating to implication and functionality such as: orientation (begetting, conclusion, interruption, introduction, suspension, trigger); stratification (background, figure, foreground); rhetoric (affirmation, announcement-reminder, call-answer, deflection, parenthesis, reiteration, sign); rhythm (pedal). Each one of these had its own symbol and Roy created a graphic score of Bayle’s Ombres Blanches using his functional symbolic lexicon.

In another collection, there were dozens of illustrations such as sketches, tables and study scores (Decroupet and Ungeheuer, 2002); sonograms (De Lio, 2002); and a beautiful “transcription and graphical rendition of a programming sequence” from Risset’s Contours (Di Scipio, 2002).

More recently, some interesting work has emerged, in particular the “Signed Listening” project which allows a listener to annotate a sound file using certain types of graphic representations (Donin, 2004), and the work of Couprie (2004) which builds on the model of the Acousmographe developed by INA-GRM in France (See http://www.ina.fr/grm/acousmaline/polychromes/index.fr.html ). Couprie has developed a graphical representation tool to produce an enriched listening experience.

6.3 Practical Analytical Methodology

As noted above, the aim of the analysis was to apply the SIAM framework to Denis Smalley's *Wind Chimes*. This was achieved through a combination of signal analysis and critical listening.

6.3.1 Digital Encoding

The CD recording was digitally transferred to computer disk at 44.1 kHz sampling rate. The original stereo version of the work is used for listening purposes and for examination of each channel's spectrum individually. In addition, a mono version of the work was created by combining the two channels to create a single spectral representation of all components of the work for further examination. The sonogram was chosen above any other form of signal analysis since it provides the most accessible way of assisting with the segmentation of simultaneous sounds. It is also important to note here that this study does not include an analysis of the spatialisation aspects of the work.

The program Adobe Audition (version 1.0) was used to both play the sound files and to create various spectral displays of frequency versus time. A "patch" was created in the graphical programming language PD (Puckette, 1996) to assist with the aural identification of significant frequencies and pitches. The patch consisted of three sinewave oscillators that could be tuned to any desired frequencies and amplitudes.

6.3.2 Segmentation and Section Formation

The work was first segmented into 30 second blocks since longer periods may become difficult to handle due to listener memory considerations. This length of time also provides a convenient screen display to work with. Four to five second windows within the 30 second blocks were used to approximate the "sliding window" of our working memory – an approach derived from the “event structure processing” theories.
The work was further segmented at obvious sonic boundaries such as those places where there may be silence or where there is a long, sustained sound. This secondary form of segmentation results in the delineation of a number of “Sections”. In essence a hierarchical form of segmentation in time has already begun. While not necessarily appropriate for all styles of acousmatic music, some of which may be through-composed or in “moment form”, the creation of sections is appropriate in the case of Wind Chimes. Nine sections were identified by such boundaries within the 15 minute work, and they have been used as the basic frame of reference in the analysis that follows, with each section referred to in a numerical way (Section 1, Section 2, etc...). These sections are not all the same duration. Some sections are quite short (eg. Section 1 is only 34 seconds), and some sections are quite long (eg. Section 5 is nearly two minutes in length). A re-evaluation of the initial sectioning does take place in the discussion of the “assimilation and meaning” part of the SIAM method.

### 6.3.3 Initial Analytical Process

Analysis proceeded in a linear way from start to finish. For each section (Section 1, Section 2, etc...), a set of observations was recorded in text form, and a pictorial representation of separated sound events was drawn in pencil on paper along a timeline. Analytical data was also drawn on the pictorial representation. In the text that follows, the terms “sound object” and “sound event” are used interchangeably since a sound object is a sound event in the context of the work’s experienced sonority. In general, the term “sound event” will be used in preference since it is a standard term used in sound synthesis where a “sound event” is said to have a certain start time, duration, and an inter-onset time before the next “sound event”.

The data that was collected included a sound event's: start time; duration; perceived pitch or significant frequency components or both; graphical indications of amplitude envelope; graphic symbols depicting special features, e.g. pitch glissando. A "discussion" passage was also written for each section in order to interpret the observations that were made.

To facilitate the discussion in a written form, a label was also given to each sound event. Rather than using some abstract code as a label (eg. Ax–425), a word or phrase was used (eg. Bowed cymbal). The word or phrase used represents either the
researcher’s estimation of a possible source for the sound, or an onomatopoeic rendering of the sound in word form. It was thought that this approach to labelling may be useful for any semantic referent analysis. It is also important to stress here that the label may not be the actual source of the sound event, since this genre of music utilizes signal processing, and the experience and aural skills of the researcher are also significant factors.

Once the observations, pencilled pictorial representation, and discussion passages had been completed, an "interactive study score" was created as a Flash interactive (See Figure 6.1). This attempted to establish a dynamic relationship between the analytical data, the spectral representations, and the audio sounds of the work itself.

6.4 Design of the Interactive Study Score

Figure 6.1 shows a typical screen from the interactive study score for Wind Chimes. The first design question requiring resolution was how to maximise the amount of data displayed while maintaining compatibility with common computer screen sizes. The interactive is designed for a monitor screen size of 1024 x 768 pixels, which was considered the best compromise.

The aim was to display aspects of the frequency spectrum (rather than waveform amplitude) versus time, display graphic symbols depicting individual sound events, display text providing information on the sound events (a label, important frequencies, pitch, special features), display start times, give some representation for duration, display other graphic symbols suggesting relevant patterns (pitch glissandi, filter changes, rhythmic patterns, etc).
Figure 6.1: Screen shot of the Wind Chimes “interactive study score”

The screen is largely divided into two panes. The top pane displays two sonogram representations of frequency versus time for the same section of the work. The sonogram placed at the top of the pane shows the full spectrum from zero to 20,000 Hertz. This is especially useful to focus on what is happening in the upper part of the spectrum above 5,000 Hertz for example. Immediately below is the same passage of the work, but displaying the lowest part of the spectrum – from 0 to 1,500 Hertz. The display of this part of the spectrum is useful for concentrating on pitch aspects of sound events.

Underneath the sonogram pane there is a larger area that is reserved for graphic symbols and text - the product of human analysis of the work. This area is meant to depict the segmentation of the work into separate sound events, and to provide some hard data on those sound events, such as important frequencies and pitches, as well as descriptive information (See Section 6.3.3 above for a description of the use of labels). A grid background was constructed to facilitate alignment with the time scale of the sonogram plots, and section labels are included in this area.
The symbols are only crudely placed vertically on the screen from lower frequency to higher frequency since some sounds may be spread broadly across the spectrum (noise) whereas others may be very pure and only contain a few frequency components. Vertical placement also depends on the amount of activity at any one time and the amount of screen real estate available, resulting in certain compromises in visual placement of graphic symbols. Horizontal placement represents fairly accurate representation of an event’s start time and approximate duration.

Just under the sonogram representation pane, and on the left hand side of the screen, lies a set of playback controls. These are used to initiate the playback of the sound track of the piece, and to control the playback in synchrony with the visual representation. There is a play/pause button, a return to start button, and a scrub-like button that permits the user to move to a particular point in the sound track. It also acts a progress bar during playback. The media controller was derived from the controller developed in the online tutorials provided by the Glasson Murray Group (2005).

A uniform duration of two minutes per screen was used to allow the comfortable depiction of sound events on the width of the screen. This length is a compromise between the macro and micro views of the work, and doesn’t represent how the work is sectioned. It is possible to zoom in to look at details using functionality available within the Flash player (right click the mouse to see available functions).

6.5 Results

In the discussion that follows, the conventions observed for the labelling of frequency, pitch and time are:

- Frequency – In Hertz (Hz).
- Pitch – Pitch class, octave, number of cents above (plus) or below (minus) the pitch class, e.g. (C4 + 23) is 23 cents above middle C.
- Time – Minutes, seconds, and hundredths of a second, e.g. 3:42.75
Still images of the interactive study score have been included as figures in the following text, but it is recommended that the reader runs the interactive study score provided on the accompanying CD-ROM, since this will allow the reader to listen to the work and to follow the study score in real time. (Instructions for installation and use are provided in Appendix 2.)

In the text that follows, the work has been divided into sections, and a set of observations and discussion is included for each. Then the whole work is reviewed.

![Figure 6.2: Wind Chimes study score 0'-2'])

6.5.1 Section 1

[0:00.00 - 0:34.60]

Observations

In the SIAM framework, segregation of sonic objects is the first task along with specification of sonic properties and factors involved in identification.

The piece begins with a single sound event, or is it? We have ambiguity straight away. The work begins with a percussive chime-like sound that lasts for four seconds, but
from the sonogram it would appear to be a combination of two sounds with an identical start time. An examination of the spectral plot reveals a spectrum with twin peaks dotted at regular intervals, e.g. 959 Hz and 1016 Hz. These frequencies correspond to the pitches (A#50) and (B50).

We would expect sounds a semi-tone apart would cause beating or roughness. There is some experience of this at the start of the sound event, but the (A#50) component is the dominant component of the two.

Over eight seconds of silence precedes the second sound event which also seems to be a compound sound event consisting of several gong-like sounds. Although this event has a fast, percussive attack time, it has a number of its spectral components extended through time for nearly 18 seconds from 0:08.62 to 0:26.40. Its pitch is estimated to be (A#48). This sound is sustained as a backdrop to other sounds that are superimposed.

The next event is a very low “choof” sound with spectral components as low as 30, 60, and 90 Hertz. It begins at 0:10.10 and lasts for 1.05 seconds. Its pitch is around (B32), although it is so spectrally dense that we should describe it as a shaped-noise sound with a pitch “centricity” of around (B32).

The low choof sound is closely followed by another event. This event is really a collection of contributing smaller events – many high tinkling chimes, and the collection acts as a texture that lasts nearly 16 seconds from 0:10.32 until 0:26.18. It is almost as if the choof sound has triggered this chime texture coming 0.2 seconds after it like a chain reaction.

Already we can see some of the binding forces coming into play in the relationships between events. Proximity in time in the form of simultaneous or near simultaneous attacks binds the events together.

The high tinkling chimes are sustained along with the sustained selected gong-like spectral components until another low choof sound is repeated at 0:17.40. Although repeated, it is a varied repetition in that a new element of frequency transformation is added - a subtle dip and rise in frequency within its dense spectral fabric.
The choof sound is repeated at 0:18.90, then again and again at decreasing inter-onset times creating an acceleration, and leading up to another loud percussive sound at 0:26.18. This sound appears to be a loudly struck piano frame. Superimposed over this sound, but at a slightly delayed onset, is a lower bass “hum”. It has two peaks in its spectrum at 240 Hz (A#+50) and 39.8 Hz (D1#+42).

**Discussion**

The work begins in a spectacular fashion! Controversially, we can see that the piece begins with, not a single sound event, not a combination of sound events, but a single sound event that is a combination of sound sources. I shall call this a “compound event” since it appears to have more that one source or cause, and yet we perceive it more as a single entity.

Is this a case of “selecting from the evidence” (schema-based segregation), or “partitioning the sensory evidence” (primitive-based segregation)? (See Chapter 4, Section 4.5.2) I suspect that close proximity in attack time is forcing a chunking of two events into one. In other words we have an example of a “fusion of the evidence”. The significance of beginning the piece with such a phenomenon cannot be underestimated. The use of such a “compound sound event” is an important symbol for the whole work.

The initial sound event’s pitch is centred around (A#+50) and it has a percussive chime-like spectrum, and that seems to be its apparent source.

When considering the dominant discourse, we would have to observe that source bonding is obvious but weak, sounds have vague sources which are related to percussion instruments, and the dominant discourse is typological-relational (syntactic).

What relationships are operating between the segmented entities? What type of syntax is being applied? Well, first of all, there is evidence of “proximity in frequency”. Many of the pitches are built around a “pitch centricity” of around (A#+50) – which provides a binding force in frequency.
Proximity in time binds separate sounds together into an experience of a single “compound” sound event by having the same attack point or nearly coincident attack point. For a nearly coincident attack point, the impression can be created that one portion of the sound is triggering the “consequent” portion of the sound event.

Added colouration is provided by superimposing secondary events at approximately a tone apart in pitch/frequency. This creates the impression of a separate sound event but as an embellishment of the primary sound event due to the proximity in frequency and time, coupled with spectral similarity.

Contrast is also used for variation. There is temporal contrast when short repeating choof sounds contrast with sustained gong sounds, and there is spectral contrast provided by the three layers of: spectrally dense (choof sounds); spectrally pure (gong sounds); and rapid repeating chimes.

6.5.2 Section 2

[0:38.22 – 1:23.25]

Observations

This section is just over 45 seconds long and is marked by three significant events. Each one is quite complex and overlaps with its neighbour.

The first significant event at 0:38.47 is a compound event, but unlike the compound sound event we saw in Section 1, this one is a chain of sub-events linked serially. It has a sharp attack that is a low “klunk-like” sound lasting about 1.33 seconds. At the same time there is a rapid upward sweep in its high frequency components with a perceived pitch glide from 1105 Hz to 1502 Hz (C#6 to F#6).

After this explosive klunk-like sweep, the spectrum is sustained for around 17 seconds until the next major event at 0:55.56. During this “continuant” phase, there is a treatment of the high frequency components that we could describe as a “melody of higher partials” between 5 kHz and 8 kHz. For example there is a three “note” figure beginning at 0:41.24 with significant frequencies at 5447, 6092, and 5655 Hz or in terms of pitch: (F8-44), (F#8+49), and (F8+21). Thus we have auditory streaming of partials that are roughly a semitone apart. Inter-onset times are 1.46 secs and 0.46 secs.
respectively. The latter two “notes” of the three note figure repeat. Added to this scenario is the coloration of extra partials at 5998 and 5908 Hz. Their pitch equivalents are (F#)+22) and (F#8-3) resulting in a certain beating effect.

On top of all this treatment there is evidence of an inverted U-shaped transformation to the frequency spectrum that results in a “smearing” of the perceived timbre. If that wasn’t enough, a sustained harmonic spectrum is faded in – beginning at around 0:44.96. It has a defined pitch of 1102 Hertz (C#6+17) and is sustained right through the next two attacks until 1:09.00.

To summarise then, this first compound sound event has three sub-components that are serially connected. It begins with a klunk-sweep from C# to F#, continues with spectral colourations centred around F#, then has a fairly pure spectrum added centring on C# that is sustained through the next event.

The second significant event begins at 0:55.56. It is another compound sound event with three components. We see a return of the low choof sound from Section 1. A second component enters at 0:55.78 and continues until 00:59.78. It is a pitch component with a frequency of 240 Hz (A#3+50). The third entity consists of sustained high frequency components with significant frequency components at 952 Hz (A#5+36) and 2013 Hz (B6+32). There are also intermittent colourations of frequencies above 5 kHz, e.g. 10,766 Hz at 1:01.00. The third entity’s sustained components last until 1:09.00 and therefore overlap with the next major event. They also form a “spectral harmony” with the continuant from the first event forming an interval of about a minor third: (C#6+17) and (A#5+56).

The third significant sound event in Section 2 begins at 1:05.17. It is the sound of many chimes in the higher register (> 800Hz). The chimes are excited repeatedly for around 18 seconds. The aggregate fades, but some frequency components ring on, e.g. 2989 Hz (F#7+17).

Discussion

In Section 2 we see the continuation and elaboration of the “compound sound event” technique where several different sound entities contribute to a complex attack, then
we have the “continuant” phase where certain spectral components are drawn out, and a pitched element is added.

The continuant components become like a connecting thread that links each sound event through the application of horizontal streaming. Smalley is creating a spectral contrapuntal technique where attacks don’t coincide but where the sustain portions of sound events continue past the attack of the next sound event.

A further binding element is the relationship between pitches. The first event combines (F# + 49) with (C# + 17). The C# overlaps with the pitch (A# + 50), with a final return to (F# + 17). These leaps represent the consonant intervals of a fifth and a third, creating a type of triadic exploration.

There are links with the first section, with the return of the low “choof” sound and the use of the (A# + 50) pitch centricity.
6.5.3 Section 3

[1:22.84 – 3:33.69]

Observations

This section can be roughly divided into three phrases that shall be called “episodes”. Episodes 1 and 2 overlap and Episodes 2 and 3 are separated by silence. Episode 3 is a varied repetition of Episode 2. This refinement to segmentation in time represents the recognition of “pattern” objects as a part of the consideration of horizontal integration (See point #5 in Table 5-1).

Episode 1 begins with a compound sound event. Here temporal proximity groups, while the spectral/temporal properties segregate. Firstly three sound events begin roughly simultaneously. The first component is a gong-like attack lasting from 1:22.37 to 1:26.66. The second component consists of a long continuous set of spectral components, sustained until 2:00.00 and then faded out (at 2:05.00). The most prominent frequencies in this sound are 133.9 Hz (C3+43), 247.36 Hz (B3+2), and 318.0 Hz (D#4+38). The sinewave generation patch was used to verify the “perceived
pitch” associated with each of these frequencies, since the sonogram was far too dense to derive any definitive conclusion regarding perceived pitches, or even distinctive frequencies. While the frequencies of the perceived pitches are not harmonically related, the intervals between them of roughly a seventh and a third do encourage a consonant interpretation.

The third component of this initial composite sound event is a series of low chime sounds that sound like a Javanese Anklung, but retaining a metallic timbre, not bamboo. These “shaken Anklung-like” sounds continue until 2:00.00 and fade out at 2:05.00.

Episode 2 begins by fading in from 1:55.40. It contains some complicated sound structures too. It begins with a cluster of sounds that appear to be processed Anklung sounds. The processing provides a “scratchy” effect through some sort of time-varying filtration, reinforcing the upper partials within the spectrum. There are lots of high frequency components within the spectrum, such as the one around 2568 Hz, reinforced at short time intervals and giving the impression of “figures of two or three events” like little motives. Fading in at a slightly later time of 2:03.00 are some amplitude modulated Anklung sounds. This granulated and filtered texture is used as a background until the end of this episode. A distinctive figure of rhythmic patterns within the scratchy sound spectra enters at 2:28.62. The episode is terminated with the sudden entry of a muffled bass drum-like sound at 2:51.00.

After the drum has died away, Episode 3 begins with the scratchy sounds using a motive we’ve just heard. The scratchy sounds continue until, at time 3:06.94, there begins an almost literal repeat of the “rhythmic patterns” within the upper partials of the scratchy sounds, accompanied by the granulated Anklung sounds in the background (as before). While this repeated section continues for over 20 seconds, a new element fades in at 3:12.00 and gradually gets louder. It is the sound of bells repeating. Their pitch becomes more distinctive with time and is centred around 806.45 Hz (G5+48). There is also a hint of some bells pitched at 664.54 Hz (E5+13), and the section actually ends with a grace note/final note combination on the (E5+13) pitch.
Discussion

So what are the operational forces within this section? On the semantic-syntactic scale, the sounds events are skewed more to the syntactic end. While there is still some source bonding, with the sound events sounding somewhat like chimes, gongs, angklungs, and bells, their spectra are stripped apart and given special treatment.

Compound sound events are created with a complex attack and then carefully crafted sustain portions extend certain preferred frequency elements that provide a fusion mechanism or “glue”. The repeating angklung sounds are used as background material throughout each episode and therefore create a unifying element, even though they are subjected to some variation treatments.

Smalley cross-fades from the almost “real-space” source-causes in Episode 1, into processed source-causes in Episode 2. The spectra are related, but they are transformed.

Distinctive timbral figures are introduced in the upper portions of spectra, thereby creating a rhythm of partials – I shall describe these as “spectral motives”.

A sudden percussive sound is used as punctuation (as we saw in Section 2), before a varied repetition of Episode 2 is heard in Episode 3.

Overall, Section 3 moves from “spectrally pitched” (around a B natural centricity) to a spectrally dense and rhythmic middle section to finally arrive on a definite pitch as its concluding event, articulating the minor third interval of G to E, in a pseudo cadence.

Sustained spectra are used to connect frequencies through time, while loud percussive sounds are used to delineate and punctuate sections and phrases.
6.5.4 Section 4

[3:33.82 – 4:49.67]

Observations

At first listening, Section 4 could seem like a simple sustained section with not much activity – like a period of stasis. Upon closer inspection, however, this section is quite deceptive in that there is a lot of activity with many different sound elements.

The section begins with the same “bell” sound that ended Section 3. This is soon followed at 3:30.36 by a “compound” sound consisting of a low bell chord-like sound with some very low frequency components quickly added. The “bell-tree chord” has significant components at these frequencies: 766 Hz (F#5+60); 1763 Hz (A6+3); 2483 Hz (D#7-4). The added low frequency components are: 149.7 Hz (D3+33); and 183.7 Hz (F#3-12). These almost triadic frequency components are sustained with varying amounts of amplitude modulation for almost the entire section. They function like a pedal point in traditional music.
A soft Anklung-like sound answers the bell-tree at 3:37.46, thus introducing yet another timbral thematic element. The term “answers” is used here in the sense that one sound closely follows another sound in time, but it implies that they are related in some way. In this case the connection is timbral and it further implies a form of horizontal integration is operating.

A simultaneous attack of two more elements begins at 3:41.41 with a subtle “clay bowl” sound struck with a perceived pitch of around 286 Hz (D4-46) and a high frequency “triplet” figure with peaks at 5670 Hz and 11,380 Hz. Each of these elements has an answering part. This “triplet pattern” has since been confirmed by Smalley (2004b) as a four-event figure, with the final event being much softer in amplitude than its neighbours. As the fourth element may be masked within the sound texture, and as we are recording observations perceived by the researcher, we shall continue to refer to the pattern as a “triplet”.

The clay bowl is answered by its repetition at a different pitch of 464 Hz (A#4-5). Pitch is difficult to decide on here (even with the sine-wave generation PD patch), and it may be a pitch an octave lower at 232 Hz (A#3-5) is more appropriate. So the clay bowl has been transposed by a minor/major third interval. The “HF Triplets” are answered by an “HF Duplet” figure at the same frequencies, but slower. The soft Anklung sounds again.

At 3:51.49 there is a double strike “clay bowl” sound with perceived pitch of 234.4 Hz (A#3+9), closely followed by a new element – a soft “ice” sound that mainly consists of widely spaced higher frequencies. Its most prominent peak is at 4156 Hz.

The bell-tree derived sustained components continue like a translucent window while another short soft Anklung sound leads into a repetition of the HF Triplet-HF Duplet figural combination. This time they are punctuated by a clay bowl strike that has been ‘smeared” in frequency (HF components boosted, LF components attenuated) and time (the attack is slower). This gesture is immediately repeated but with more variation – a downward glissando of frequency components.
Then comes a familiar sound at 4:02.49. It is the very first sound of the entire work, but transposed down a minor sixth. It has “twin peaks” at 617 Hz (D$\#5$-14) and 668.2 Hz (E$5+23$).

The low sustain sounds continue, a high “ice” sound enters, HF Triplets repeat, a gap follows, then two variations of the high ice sound follow. These sounds are almost a cross-over between the high ice sound and an extremely filtered version of a clay bowl sound. A timbral linkage is being formed here.

The two sounds “announce” a return of a very dramatic sound – the original sound of the beginning of the piece at its original pitch.

This distinctive moment is followed by a repetition of the HF Triplet/Duplet combination but cross-faded with a subtle and soft “rubbed glass-rim” sustained sound reminiscent of the high ice sound. The “high ice” sound repeats, more glass rim sound is heard, the HF Triplet returns, high ice, then dramatically a percussive piano cluster enters and is then muted in a sforzando fashion. Its decay is as dramatic as its attack, and it is quickly followed by a compound sound consisting of a piano cluster with the addition of perhaps the low bell-tree sound from the start of the section. A combination of the bell decay spectrum and clay bowl spectrum lingers on in a long decay. The section ends.

**Discussion**

In this section, typological-relational discourse is dominant. Smalley makes use of the most abstract aspects of sound we’ve experience so far within the work. He delves inside the spectrum of a sound to:

- reinforce certain frequencies, for example the peaks in the high frequency (HF) triplet figure.

- extend certain frequencies, eg. The low “bell-tree chord” sound and added low frequency components.
Syntactic relationships used:

• Temporal proximity:
  
  o Simultaneous attacks create fusion between sounds into “compound” timbres, eg. “low bell-tree chord” sound and low frequency components [03:36.36].
  
  o Attacks with close proximity where one sound seems to “trigger” another sound giving the illusion of a common cause or causal relationship, eg. Two HF ice sounds seem to trigger the literal repeat of the work’s opening sound at time [04:15.74]

• Repeating entities:
  
  o The HF triplets and duplets form a “spectral motive” that is repeated regularly in varied form.
  
  o The relationship between the triplet figure and other entities is also repeated but varied, eg. The initial temporal relationships between the HF triplets and the two “clay bowl” sounds beginning at 03:41.41 is repeated beginning at 03: 55.44 with the two “clay bowl” sounds having been transformed in frequency and time.

• Timbral linkages:
  
  o There is a distinctive aesthetic texture that binds the section. An analyst trained in the Western art music tradition might call this texture a harmonic field since there are a number of significant sustained pitches. However, because this frequency interest has been pushed much higher in the spectrum, a more apt term might be “spectral harmony”. This is achieved by:

    ▪ The use of the long sustained “low bell-tree chord” spectrum, combined with a certain “pitch centricity”. A# is once again an important pedal tone with the additional use of D#.
There are timbral segregations that act as figure and ground. Borrowing from the musical vocabulary again, we might say that there is an active “timbral counterpoint” against the “spectral harmony” background. The hybrid nature of the phrases coined here indicates blurring of the separate realms of timbre and pitch, and in fact Smalley would consider this an expression of a timbre-pitch continuum.

- Colouration is used with “double peaks” added to the foundation A# and D# pedals.

- Sectional punctuation:

  - Is provided by dramatic percussive sounds like the initial bell sound and the concluding piano cluster/bell-tree sounds.

  - The opening sound and its transposed version are used as scaffolding at strategic points in the section – about half-way through and two thirds of the way through.

The above discussion leaves only one category of sounds unaccounted for. The “soft angklung” sounds are not only colouristic echoes sounding in response to the “feature” sounds early on, but also echoes of sounds occurring in the previous section. Thus they provide a binding element to previous material, just as the section’s initial bell sound was a repetition of the concluding sound of the previous section.

### 6.5.5 Section 5

[4:49.45 – 6:37.31]

**Observations**

This section overlaps with the previous section via the decay of the bell/clay bowl sound interrupted by a sharp percussive attack. The attack marks the beginning of another “compound sound”. It begins at 4:49.54 with a “struck piano frame-like” sound. This provides the attack of the compound sound. The sustain part of the compound sound is a band-limited noise-like sound with some prominent peaks or bands in its spectrum. The most prominent peak is at 830 Hertz, giving the noise band...
a perceived pitch of (G#5-1). There is some amplitude modulation, and it continues to waver in amplitude until, although partly masked by other sounds at 4:59.08, it seems to share spectral components of a single metallic anklung-like sound at 5:00.36 – with prominent components at 830 Hz (G#5-1), 685 Hz (E5+66), and 1205 Hz (D6+44). These components could be said to resemble the fundamental frequencies of a sharpened E minor/major seventh chord.

Also part of the initial compound sound is a complex sound with duration from 4:49.54 to 4:51.65. It is made up of several low, short sounds (“thuds”) at around 92 Hz and very high frequency components (3180 Hz, 4938 Hz, 7124 Hz). The pitch is non-specific and the effect is colouristic. The double thuds are repeated between 4:52.55 and 4:53.55.

Entering at 4:53.78 is a low “gong-like” sound with a gentle attack. It has prominent components at 188 Hz and 232 Hz which give it a chordal function like a dyad with pitches F#3 and (A#3-8), although the emphasis is on the F#. The gong-like sound fades out at 5:05.05.

Around 4:58.80 we have another complex event, or rather it is more like a point of accumulation of a number of events creating the impression of a focal point (see the discussion section below). It begins with a “triplet figure” of the repeated high frequency peaks we heard earlier, centred around 7124 Hz. Coming in just after the second triplet is a broadband noise sound which seems like a cross between a vocal “sh” sound and a rushing noise (traffic or jet noise) with slightly descending frequency components. It begins at 4:59.47 and lasts until 5:00.42. Add to this mix an ultra-high frequency duplet figure centred around 10,370 Hz from time 4:59.36 to 5:00.70. If we recall that there is also a metallic anklung sound at 5:00.36 then we have quite a clustering at this point.

Now from 5:01.48 to 5:02.60 there is a sound that appears to be the metallic anklung sound that has been highly processed creating lots of high frequency components that are rapidly amplitude modulated like a frenetic tremolo. Following at 5:02.60 is some sort of emphasised extension of this amplitude-modulated, filtered treatment, with sustained spectral components at 273 Hz (C#4-24) and 332 Hz (E4+11) producing a
minor third dyad effect. There are also spikes at higher frequencies like 10,380 Hz which are varied to create the “wow” effect. This two second section is also punctuated by roughly three “anklung/bell” sounds with filtered treatments.

There is incredibly rich detail here, like the low bass figure ascending from 132 Hz (C3+16) to 247 Hz (B3) at time 5:05.39 to 5:08.18. The B tonality frequencies (247 Hz and 122 Hz) are sustained through until around 5:17.11. Coinciding with the sustained low frequencies there is firstly an anklung figure followed by a repetition of the high frequency “quadruplet” at 10,300 Hz, then more anklung sounds.

At time 5:17.22, new material is introduced by another compound sound. Its attack is a plucked string with an upward glissando in pitch, closely followed by a noise component that has a downward glissando of spectral components. The noise is concatenated with a plucked bass string sound which also has its own complicated trajectory. All of this happens within the space of 1.44 seconds.

The plucked bass string is extended into a sustained sound that is reminiscent of a loose-wire, bass piano string jangling against a metal piano frame. This sound is sustained, with variations in timbre and amplitude, until it fades out at around 5:38.80. Meanwhile there is a re-appearance of the “low choof” sound that first appeared in Section 1. It enters at 5:20.90 and is repeated at ever-decreasing time intervals in an “acceleration” of time towards the next significant percussive event at 5:35.74.

Entering by stealth, a “whispering” glass-rim, rubbing-type sound begins at around 5:23.11. It appears to be a transposition of the same sound used in the previous section. It is sustained for around 17 seconds and it has widespread frequency components extending into the upper frequency register: 1355, 2865, 3911, 5615, 6428, 7861, 8520, 9604, 10727, 13361, 15220, 17505, and 19170 Hz.

There is a perceived pitch centricity associated with the sustained low string buzz sound, with its significant frequency components of 85 Hz (E2+54) and 681 Hz (E5+56), as determined by the PD-based oscillators playing simultaneously with the Wind Chimes passage. This tuning-in was necessary as an examination of the spectrogram was inconclusive. An additional 1355 Hz component from the
“whispering glass rim” sound was determined, thereby confirming a perceived pitch of between E natural and F natural.

The pitch of the low choof sound wavers. It has an initial upward glide, but when it settles on a value that value is about 85.3 Hz (E2+60). The low choof sound phrase, begins as it does in Section 1, that is accelerating towards a single percussive attack of the sound following, however it doesn’t resolve in this way at all. Instead its resolution is interrupted by an extremely dense section beginning at 5:35.52 and lasting until 5:47.13.

This dense phrase sees almost the whole frequency spectrum ablaze. It’s as though the spectra of many sounds have been smeared. The multitude of high frequency components experiences downward glissandi resulting in a tumult of whistling noise coupled with amplitude modulated treatments (tremolo). The dense phrase begins with a “four note figure” defined by a “melotron-like” sound, with heavy upper frequency processing. The pitches perceived are around 141 Hz (C#3+29), 209 Hz (G#3+11), repeated, and 105 Hz (G#2+19).

After this announcement comes a dense and elaborate passage in which it appears that the punctuating event at its conclusion is actually the source material for the middle of the phrase. The phrase’s first sound has been transformed by filtering, amplitude modulation, and multiplication to become a thick-textured, smeared layer. This phrase’s final, punctuating sound, at 5:47.07, is like the attack of a struck piano string, but with a buzzing sustain portion added. Although its spectrum is dense, there are many regular peaks like a “gnarly” pulse wave with lots of spikes. Its pitch is ambiguous in that there appear to be two pitch components centred on 106 Hz (G#2+60) and 163.6 Hz (E3-12).

The “buzzing dyad” is elongated into a sustained spectrum. This expansion or continuation is soon joined by a repeat of the high frequency “whispering glass rim” sound, but at a slightly softer level (5:48.36). Another layer is added at 5:50.85. It is a set of “ghost components” from a high cymbal sound that will strike later on. This “pre-echo” has a slow attack and enters with a couplet-like gesture, followed by a modified couplet at 5:57.50. The same combination is repeated at 6:07.96 and
6:15.16. This combination has occurred before – as a colouration back at 5:21.59 and 5:28.34, representing only the slightest hint of a significant event to come.

To recap, we have a complex attack at 5:47.04, followed by a sustain segment with colouristic layers that represent new material (the “buzzing dyad”), and selected sounds we have heard earlier. One of these sounds consists of a modified spectrum of the next attack sound – a “high cymbal”.

The “high cymbal” attack comes at 6:04.47, and it is another instance of a compound sound in that the attack also has a continuant of high chimes for decoration and selected frequency components sustained for over 15 seconds. The most pronounced of the components is the frequency 923 Hz (B5-17).

In this rich sustained portion, the buzzing dyad has (E3-12) as a significant pitch with another pitch about a sixth lower for colouration (G#+36). The other significant sustained component (B5-17) has a perfect fifth-type relationship to the (E3-12). The “high cymbal” is like adding a seventh to this pseudo-chord, and its “ghost components” reinforce the seventh chord aspect. This may be an example of a combined spectral/tonal language operating here, but this issue will be developed in the discussion section below.

Punctuating the exotic sustain is a very significant return, at 6:07.17, of the original percussive chime from the very start of the piece. There is also a hint (6:12.58) of the anklung-like sound used earlier.

By 6:21.00, only the now-fading buzzing dyad is sounding, but then some gentle “tinkling chimes” sound at 6:21.13 and they are subsequently cross-faded with the slowly expanding entry of what is termed a “glass rim chordal chorale” in the interactive study score. This “chordal chorale” of sounds related to the “glass rim” type sound world seems to have an opening statement followed by a closing statement that could be its retrograde – a mirror image in the glass.

Different pitches enter and subside in the manner of a traditional “vocal chorale”, but the frequencies are fairly high ones, compared to the range of the voice. The significant pitches are 504.6 Hz (B4+39), coupled with 1555 Hz (G6-14), and others
such as 786 Hz (G5+4), 1005 Hz (B5+30), AND 1106 Hz (C#6-4). The E tonality has given way to a B tonal centre with G used for harmonic colouration (a third apart) and C# used as a stepwise neighbour to the B. Although it only lasts 14 seconds, the “spectral chorale” concludes the section in a distinctive way.

**Discussion**

Section 5 lasts less than two minutes and yet it is extremely dense and complex. What our perceptual systems can assimilate in fractions of a second has taken a very long time to deconstruct through conscious analytical effort.

A multitude of frequency components at any one time has made pitch detection of constituent sounds from spectral components extremely difficult since there are a lot of possible sources for any one component present. We have used not only our ears but tuned sinewave oscillators to assist with the determination of significant pitches and relevant frequency components.

Smalley has continued the trend to more elaborate sound treatments and sound combinations in this section. Sound has become further abstracted from the real world referent and its application much more dense.

In addition to “sound treatment” we have observed certain syntactic traits emerging: Compound sound events; tonal centres and harmonic fields; organisation in time.

*Compound sound events*

Smalley has continued his use of compound sound events where the technique may consist of:

1. The attack portion of a sound event (eg. 4:49.54) contains a number of components drawn from different sound sources. This attack is sudden and percussive.

2. The sustain portion of a compound sound event draws out selected frequency components, from the attack, that contribute to a certain “tonal centricity” (see below). Other sound entities contribute to the “compound sustain” to add colour, variety and complexity.
3. A variation of the “compound attack” is the “point of attraction” (eg. Around 4:58.80). Sound elements are accumulated around the one point in time, often creating a complex gesture that may have no counterpart in the real world. The “compound gesture” becomes a new creation in Smalley’s musical world. A compound gesture (or “point of attraction”) can use a large proportion of the frequency spectrum. Smalley avoids a crowded, muddied sound through a skilful selection of contributing sounds, so that each one occupies its own segment of the spectrum in a way that doesn’t overlap too much with others. For example: A middle-band “sh” sound may be combined with some enhanced upper frequencies (> 10,000 Hz) and added to a low “thud” which consists of lower frequencies only. Where frequency components do coincide, these become significant frequencies that can contribute to a tonal framework (see “tonal centres” below).

Tonal centres and harmonic fields

Smalley has moved the manipulation of pitch upwards within the spectra of sounds used. Rather than creating melodies that consist of notes with definite pitches, Smalley emphasises certain frequencies within the spectra of the sounds he employs, or he transposes sounds so that they gravitate to certain tonal centres.

These tonal centres act like pedal points in tonal instrumental music. Early in the section, one such tonal centre is a quarter tone above E natural. Smalley creates a “harmonic field” using the pitches associated with an E major/minor seventh chord.

A B natural tonal centre is alternated with the E natural tonal centre throughout the section. Smalley colours these with thirds, fifths, and sevenths to create his harmonic fields.

Smalley has also re-interpreted traditional harmony and voice-leading practices by taking the principle of “common tones” between two successive chords in tonal harmony and applying it as a principle of “common partials” between two successive sound events in acousmatic music. The common partials then become a binding agent between the sounds through a strong horizontal steaming tendency.
An extension of this principle is the principle of “stepwise motion” used in harmonic progression. Smalley creates a stepwise motion of sustained partials when some dominant partials are changed to values that are close in frequency.

A beautiful application of these principles is evident at the end of the section where we have a “spectral chorale”. The B natural tonality is embellished by a G harmony with motion to B’s neighbour C# and back again.

Organisation in Time

In this section there is no strong pulse, no meter, and no rhythmic motives present, however there is careful structuring in time. The following temporal elements are apparent:

1. The attack-continuant paradigm is a strong structuring force.

2. Pseudo-rhythmic figures are used within the attack-continuant framework. Some examples are the high frequency couplets and triplets beginning at 4:58.80, or the low frequency figure at 5:05.39.

3. Sustained partials are overlapped with the next attack-continuant expansion, providing a “proximity in time” Gestalt throughout, which binds elements within the section.

4. As shown above, simultaneous attacks create “compound sounds” by virtue of their “vertical integration” or grouping. These accentuated attack points can function as a significant point of punctuation to announce a new phrase or section.

5. Nearly-simultaneous attacks create “compound gestures” that are so unique they generate interest and variety.

6. Repetition is used in a number of ways:
   a. Repeating a sound at short, and ever-decreasing, time intervals produces a feeling of acceleration, propelling the listener towards a goal event.
b. A more subtle technique, used over a wider time-frame, is the creation of a “pre-echo” or anticipation of a sound. Smalley achieves this by taking its spectrum, transforming it in some way, then placing the processed sound before the unprocessed sound in time. For example the “ghost components” at 5:21.59 are from the high cymbal at 6:04.47. This is much more subtle and effective than using the reverse order, that is playing the original sound followed by its transformations.

c. Repetition of sound in a new context is also extensively used, for example the “whispering glass rim” sounds at 5:23.80 are repeated at 5:48.36.

d. Distinctive sounds make judicious return appearances throughout the piece. For example, the return of the original percussive chime from the start of the piece returns at time 6:07.17. This is disguised with a busy phrase, but it is like a preparation of a full, unaccompanied return in the next section.

Figure 6.5: Wind Chimes study score 6'-8'
6.5.6 Section 6

[6:41.33-7:53.88]

Observations

The beginning of Section 6 is a very distinctive moment in the work. The first significant length of silence in the piece, approximately 7 seconds, is followed by a reprise of the opening sound of the work. It comes at 6:42.26 but it is the original transposed up by 45 cents or around a quarter of a tone. This sound provides quite a distinctive beginning to the section, and one frequency component rings on – 1987 Hz (B6+9).

At 6:43.84, there is a varied repetition of the “glass rim chordal chorale” we heard at the conclusion of Section 5. Its significant frequency components are 502 Hz (B4+27) and 3259 Hz (g&+66). Instead of this gesture fading, as it did previously, other components are added in layers to build the texture even more. A denser spectrum enters with a slow attack from around 6:48.79. It has many very high frequency components, and a dominant one is 2217 Hz (C#7). It is like a stepwise progress from the B natural sonority. Soon another layer is added with a slow attack. This layer has a vocal-like or trumpet-like spectrum. It is more harmonic but with less high frequency components than sounds heard recently. Its pitch centricity seems to hover around 202 Hz (B6+39), reinforcing the B natural sound world.

Overlapping the fade-out of these layers is the sudden sound of a plucked bass string at 7:09.22. It has string frequency components of 83.6 Hz (E2+25) and 498.6 Hz (B4+16). This sound is reminiscent of the plucked bass string in the last section with a similar pitch. Its pitch is centred on the (E2+25) with the (B4+16) a likely sixth harmonic. Other peaks are at 416 Hz (fifth harmonic), 332 Hz (fourth harmonic), 249 Hz (third harmonic), and 167.5 Hz (second harmonic).

Almost immediately several sounds follow, as if triggered by the pluck bass sound: a short anklung-like sound at 7:10.10 with pitch 401.7 Hz (D4+42); three couplets of high frequency components (10,104 Hz and 11,347 Hz) beginning at 7:10.71; a bass vocal-like sound at 7:10.90 with significant frequency component at 480.5 Hz (B4-
48). This lasts for several seconds and is a detuned “response” to the bass string’s “call”.

At 7:12.71, the accelerating low choof sounds make a re-appearance, but this time several iterations of the gesture are mixed over each other in a fugal cascade.

Most of the sounds from this phrase are repeated with variation: the bass string is repeated with a slower attack at 7:13.60, and its timbre is almost harpsichord-like; more high frequency couplets are repeated at about 7:17, 7:21, 7:23.50, and 7:27.50.

A tremolo "seventh component" makes an appearance at 7:14.04. It has a soft attack and is described as a "seventh" as its pitch is 293.6 Hz (D4) and represents a seventh above the prevailing E tonality provided by the plucked string sound. This seventh component helps to fill the texture at 7:22.17 and at 7:25.85 we hear it combined with a low sustained sound.

After all of this frenetic activity the low sustained sound entering at 7:25.85 lasts until about 7:45.00. It has strong frequency components and 81 Hz (D#2+70), 156 Hz (D#3) and 190 Hz (F#3+46). Sometimes pitch is difficult to ascertain but it seems detuned slightly (by 55 cents) from the (E2+25) of the bass string like sound heard earlier.

Over the top of this sustained sound, soft bell sounds are added - a group of four repetitions beginning at 7:26.24 and a group of three beginning at 7:32.85.

Coinciding with the third bell sound (in the second group) a scraping broadband sound enters and adopts the same rhythm, that is it repeats at about one second intervals but on its third repeat, it varies with a louder double attack. Upon the second of the attacks, a modified bell sound is superimposed. This double attack noise gesture with bell is repeated three times, and as the bell sound gets softer, the final iteration of the noise band has an anklung sound imposed.

The noise bands are accompanied by filter sweeps in a variety of directions: firstly an upward sweep, then up-and-down, then a U-shaped movement, then upward and shallower.
Just as this idea is losing momentum, another complex passage begins its attack at 7:44.66.

From 7:44.53 to around 7:51.65 there is a very dense texture which seems to serve as a transition to the next section. At 7:44.53 there is a compound sound made up of: the scraping noise (7:44.53 to 7:45.51), bell/chime sound at 7:45.01, a very high frequency couplet distorted, a two note figure with pitches (D#5+32) and (D5+31) from 7:44.53 to 7:46.51, the two note figure is repeated and then just the first note (D#5+32) sounds from 7:49.10 to 7:50.42. There is also the tremolo seventh component sounding from 7:46.18 to about 7:49.00 with pitch D4.

At around 7:51.35, a gong like sustained sound enters and marks the start of a new section.

**Discussion**

This section marks the start of the second half of the entire piece through the distinctive use of the opening sound (surrounded by silence).

Smalley then re-uses a sound that we’ve just heard, but he extends it by sustaining it and adding crescendo. He is extending the “spectral chorale” idea. This “spectral chorale” uses vocal-like, or trumpet-like spectra, introduced with a pitch centricity around B natural, with a “neighbour note” of C#. It builds to the point where the ear is led to an explosion which is the centre-piece of the section.

This mid-portion of the section is spectrally dense with the re-use of some previous material, but in altered forms. We hear a plucked string-like bass sound at 7:09.22, but it is different to the one we heard at 5:17.22 in Section 5. The accelerating “low choof” sounds, from the opening section, return at 7:12.71, but they are mixed over each other in a type of fugal presentation. The middle section has a prevailing E natural pitch centricity but Smalley adds colourations at around a whole tone or a semi-tone away from this pitch centricity. For example the bass string and the bass vocal sounds have strong components of 498.6 Hz (B4+16) and 480.5 Hz (A#4+52) at 7:10.90.
We also hear a tremolo “seventh component” sustained sound repeated through this mid-portion of the section. I’ve called it a “seventh component” since its pitch (D4) is an interval of a seventh above the prevailing pitch centricity of E natural.

The high frequency couplet/triplet sounds also make a re-appearance in this mid-portion of the section.

After a very active mid-portion the final third of the section ebbs to a sustained portion punctuated by a regular repetition single bell sounds or single chime sounds. The sustained sounds are very low and support the underlying E natural pitch centricity, although slightly detuned from before.

The introduction of some broadband sounds with filter sweeps heralds the transition to a phrase that links to a new section. The link phrase has echoes and transformations of previous material.

In Section 6, we can observe that Smalley continues to use the following:

1. Expansion of the “spectral chorale” idea.

2. Reuse of previous materials transformed and mixed into different relationships.

3. The use of compound sound objects made up of a number of separate sound events.

4. The establishment of a “pitch centricity” used to bring together materials within certain phrases.

5. Colouration of the basic pitch centricity through use of subsidiary sounds a semitone or a seventh apart from the prevailing pitch centre. (I am avoiding using the term tonality.)

6. Use of all parts of the spectrum for structural purposes.

7. Organisation of time has:
a. short-term components - for example certain rhythmic elements like the accelerating choofs or regular bell sounds.

b. longer term components - we hear an explosion of sound activity followed by long, sustained sounds which have a much less dense frequency spectrum creating a certain breathing pattern in the phrasing. The pattern is like expansion followed by relaxation.

![Wind Chimes study score 8'-10'](image)

Figure 6.6: Wind Chimes study score 8'-10'

6.5.7 Section 7

[7:51.35-10:18.57]

Observations

What is immediately apparent about this section upon first view is the striking patterns and shapes on the spectral plot. After about 9:30.00, there are oblique lines ascending and descending, denoting some forms of variable bandpass filtering.
The section begins with a gong like sound which is sustained through most of the section, up until about 9:50.00. Its amplitude rises and falls periodically throughout at intervals of between five and 10 seconds. It has two prominent frequencies of 78.69 Hz (D#2+20) and 151.02 Hz (D3+48) and there is a slight beating between these frequencies.

At 7:55.82, another sustained component enters which is higher and it wavers in and out in amplitude at about twice the rate of the gong like sound. This component’s main frequency is 668 Hz (E5+23) and it lasts about the same length of time as the gong like sound. It provides added colouration. With the gong sounds providing a sustained foundation, there is other activity added at 7:54.60 with a scraped metal type sound which is then subjected to quite rapid amplitude modulation lasting till about 8:02.00. Superimposed on this is a familiar pattern from previous sections - a very high frequency triplet rhythm followed by two duplets in quick succession. This pattern has prominent components at 5600 and 11,500 Hz (like before).

Just when the triplet-duplet pattern finishes, we hear a compound sound event. At 8:04.21, a short scraping noise sound is heard lasting for 1.22 seconds. It appears to be the same sound heard in the previous section at 7:35.01, but without the additional sweeping filter treatments heard earlier. In the middle of this attack of noise, another gong is struck at 8:04.81 and dies out over 18 seconds. It has prominent frequency components at 466, 656, 863, 1216, and 1592 Hz. There is some pitch ambiguity here and it is hard for the listener to isolate a specific pitch experience. With the aid of some additional sinewave oscillators from our PD patch, we can say that the prominent 656 Hz (E5-9) first dominates and then the 466 Hz (A#4) component is sustained more. The 656 Hz component is also coincident with the other sustained component, highlighted earlier and still ringing at 668 Hz (E5+23). There seems to be a pitch centricity built around a cluster of D, D#, and E.

At 8:11.06, the high frequency triplet-duplet combination is repeated. Then the glass rim type sound, with many components, fades in at 8:15.48. It gives us another glimpse of the chorale like passage referred to in the previous section then fades out by 8: 25.70. Beginning at 8:20.00, there is a long sustained sound that lasts until about 9:27.27. It adds yet another layer to the texture of the section. It is a scrabbling,
low sound within a band of frequencies of 200 to 1600 Hz. It is periodically filtered leaving U-shaped traces on the spectral plot in about seven places.

At 8:25.70, we hear another sound making a reappearance from an earlier section, it is a plucked string with an upward glissando. Its pitch begins at 820 Hz. Almost coincidentally we hear the high frequency triplets. A short while later, at 8:33.16, we hear the same plucked string with glissando, but transposed down a perfect fourth (608 Hz, D#5-39). A very low bass sound is added as a “continuant” with components at 55.8 Hz (A1+25) and 50.5 Hz (G1+51). The two frequencies create beats of about 5 Hz. Answering the low plucked string sound is the high frequency triplet-duplet combination again.

What follows this short active phrase is an extended passage which is built upon the sustained base of the low gong (D#2+20 and D3+48), the higher sustained component (E5+23), and the low scrabbling sounds (200 to 1600 Hz). Added to this base are the high frequency duplet-triplet combinations repeating periodically. Another set of sounds are added higher in the spectrum too. There are two timbres: a bowed metal sound and a bowed glass (or rubbed glass-rim) sound. At first these sounds are introduced, at 8:47.80, as a “response” to the high frequency triplets “call” at 8:46.88. Then the metal and glass sounds create a rhythmic counterpoint with the repetitions of the high frequency triplet-duplet combinations. The metal and glass sounds are very high and so pitch is not of primary concern here. For the bowed metal sound it has a prominent frequency at 841 Hz (G#5+22). The bowed glass sounds have prominent frequencies of 1650 Hz, 3300 Hz. There is some treatment of these sounds too. There is a glissando of the bowed metal sound at 9:10.81 and there is a pitch glide down to 1605 Hz (G6+40) and up again in the final bowed glass triplet at 9:20.86.

The beginning of the next phase is punctuated by a sudden foot stomp-like or bass drum-like sound at 9:26.81. It marks the introduction of about a 30 second passage that combines the treatment of source material (the scrabbling sounds?) by what appears to be 11 different bandpass filters. Seven of the filters have centre frequencies that are ascending, (for example 1158 Hz to 3118 Hz and 15,412 to 17,773 Hz) and four have centre frequencies that descend over the 30 second period (for example 7973 Hz to 3029 Hz). The result is a spectacular display of opposing
timbral motions. The gong sound and the higher sustained component continue to underpin this phrase until they begin to fade at about 9:50.00. The high frequency triplet is repeated at 9:39.39 and is answered by another foot stomp at 9:40.46.

The final high frequency triplet at 9:46.78 signals the introduction of two new elements. At 9:45.30 a sustained passage of bandpass or comb-filtered noise enters and lasts until 10:17.44. The filters have bands about 600 Hz in width and the filtering is somehow modulated in an exotic fashion - first slowly, then rapidly, then slowly, then rapidly.

At around 9:50, we have filtering at the other end of the spectrum. Low frequencies (< 300 Hz) are filtered in a stepwise way creating ascending, then descending, melodies.

While all of this exotic filtering is going on, a contrast is added with chimes ringing through from 9:58.33 until the end of this section at 10:18.57. The chimes have prominent frequencies of 318 Hz(D#4+38), 658 Hz (D#5+96), 920 Hz (A#5-23), and 2252 Hz (C37+27).

**Discussion**

The organisation of this section seems to be in overlapping phrases built upon an extended to pedal bass (gong, plus higher sustained component, plus scrabbling sounds).

The first phrase is active, with modulated scraped metal sounds, the high frequency triplet-duplet components, scraping noise, and the second gong sound. The second phase is the linking phrase consisting of the high, glass-rim chorale section bridging to a short third phrase of plucked string activity.

The fourth phrase is an extended passage that includes a rhythmic counterpoint of sounds in the upper parts of the spectrum. The high frequency triplets-duplets create a call-response and rhythmic interplay with the bowed metal, rubbed/bowed glass sounds. Smalley carefully controls use of the full spectral range with lower sounds creating a long sustained underpinning and a pitch centricity of D/D#/E. Sounds in
the middle of the spectrum create timbral interest, through the use of extensive filtering, and higher frequency sounds are used to create rhythmic interest.

The fifth phrase is marked by a sudden foot stomp sound and is characterised by different elaborate filtering treatments afforded to the differing spectrum registers noted in the previous phrase. The chimes introduced at the end of the section reinforce the D#/E centricity.

In Section 7 we can also see continued use of previous structural devices: reuse of sounds from earlier in the work; the use of pitch centricity (although much less prominent); the use of compound events. What is different about this section is the increasing use of signal processing which reinforces the movement away from source-cause real space, towards a typological-relational music space.

Figure 6.7: Wind Chimes study score 10'-12'

The Development of a Cognitive Framework for the Analysis of Acousmatic Music 178
6.5.8 Section 8

[10:18.57-12:50.00]

Observations

This section begins with the chimes ringing through from the previous section at the same frequencies. Superimposed are a series of short chirping sounds at high frequencies (4500 Hz, 5211 Hz, 5400 Hz, 6100 Hz 4900 Hz, etc) creating a pseudo melodic line. At 10:31.92, filtered low-frequency sounds create another pseudo melody - ascending. It is a repetition of what we heard in the latter part of the previous section at 9:50.00.

"Bird whistles" then enter at 10:40.00, occupying the frequency range 1200 to 3000 Hz, and lasting for about seven seconds. A new “phrase” begins at 10:47.82 that is an interplay between three sets of sounds using the same three-pitch motive, but in different frequency registers. The first sound type is that of a “granulated high frequency triplet type” which uses varying amplitude modulation in the granulation. It repeats the three-pitch motive with frequencies 1383 Hz (F6-16), 1628 Hz (G#6-34), and 1958 Hz (B6-15). Within the same frequency range but creating a contrapuntal interplay are some "glass-rim" figures using the pitches 1630 Hz (G#6-33) and 1945 Hz (B6-26). Closely following the glass-rim sounds at time 10:50.79 an organ-like three-pitch motive enters with each of the pitches entering sequentially at 345 Hz (F4-18), 411 Hz (G34-18), and 490 Hz (B4-12). Each of the pitches is then held to create a chord (like a diminished triad). The three sound types work with each other in a call-response fashion.

A new phrase begins with a “punctuating sound” – a short, loud, scraping noise sound at 11:05.05. This noise sound functions as the attack for a long sustained sound beginning at 11:05.68. This compound sustained sound is a complex mixture of low frequencies, high frequencies, some organ like timbres, the high frequency triplet idea, and they are all treated with amplitude modulation. This complex texture lasts about a minute and 20 seconds and the various constituents and treatments are varied over its total duration. It has some pitch centricity gravitating to C#, but there is an interval of a seventh below this that sounds and creates some beating effects.
Punctuating this elaborate drone, new elements are injected with some violence at 11:20.87 where we hear a loud “struck piano frame” as a compound sound event - combined with a plucked string with simultaneous glissandi up-and-down, combined with a high frequency duplet. A varied “struck piano frame” sound answers quickly at 11:22.28.

More complex drone, then the sound of a “struck saw” with an upward glissando in pitch (up from 568 Hz, D5-2) is heard. Another scraping noise follows at 11:32.31 and it seems to trigger some chimes, at 11:33.09, which utilise pitches in the mid-frequency range. Their pitches are 785 Hz (G5) and 951 Hz (A#5+35). The chimes continue for about 14 seconds and then only the complex drone texture is left.

At 11:54.71, high chimes enter and ring with pitches 2906 Hz (F#7-31), 3662 Hz (A#7-31), and 4380 Hz (C#8-21). They ring on for about 14 seconds, and while they do, the complex drone texture subsides then re-emerges. Another percussive stroke of punctuation occurs at 12:07.70 where another struck piano frame sound “triggers” a continuant that is like a Wurlitzer organ chordal sound. This sound is sustained but treated with amplitude modulation that varies from a very fast to slow, and then to moderately fast over its 12 second duration. The “chord” uses pitches which include 567 Hz (C#5+40), 714 Hz (F5+48), and 1131 Hz (C#6+34). The complex drone texture begins to die, but has some last death throes at 12:20.05 and 12:23.60.

At 12:23.55 the Wurlitzer organ sound enters again, but this time it is much richer with more frequency components and pitches. The pitches have staggered entries and the predominant ones are 546 Hz (C#5-24 = C5+76), 673 Hz (E5+37), and 868 Hz (G#5+77). There is amplitude modulation used as a prominent treatment here too. It varies from fast to slow to moderately fast. Vibrato (or frequency modulation) further blurs the texture and it increases over the 26 second duration of this sound.

The rich organ like texture is garnished with some high frequency embellishments. At 12:38.26 we hear the reappearance of the bowed metal, bowed glass combination. At 12:42.80, a three-note bowed glass figure is answered by an elongated three-note bowed metal figure, which is quickly punctuated by a single bowed glass event, and then the section subsides.
Discussion

A characteristic of this section is that it progresses through a succession of phrases. The first phrase is a continuation of the chimes from the previous section. They are further embellished with chirping sounds and bird whistle sounds, while the bass register and is filled with the repetition of the figure recently heard in the previous section. This continues Smalley’s trend to reuse materials, both literally and in transformed forms, in subsequent passages within the work.

What is made obvious from the structure of this first phrase is how Smalley uses his materials in different registers of the frequency spectrum. This gives the sounds clarity in that they don't mask, obscure, or interfere with each other, or create a muddy texture. It also enables Smalley to create simultaneous melodies very high in the frequency range and very low in the frequency range.

The second phase, beginning at the granulated high-frequency triplets (10:47.82), develops this use of different motives in different registers, but they move closer together as the lower sounds move up to the organ-like three pitch motive and reinforce the pitch material used in the upper frequency register by the granulated high-frequency triplets and the glass-rim sounds. Once again we find a “spectral counterpoint” operating between these two upper parts - circulating rhythmically through this “diminished triad” of F, G#, and B.

The third phrase is initiated by another sound heard early on in the previous section - a short scraping noise. Smalley uses loud, short, percussive sounds to “trigger” continuous sounds. In this case the triggered sustain sound is a complex mixture of some of the individual elements just heard, for example organ sounds, high-frequency triplets, low and high frequencies, but these are subjected to elaborate amplitude modulation processing (granulation).

Such processing confirms the trend in the work of more signal transformation as the work progresses. Here we are nearing the end of the work and we are witnessing the largest amount of processing. This is quite a long phrase lasting over a minute from 11:05.68 to 12:07.70. Its underpinning is a complex drone texture, and it is punctuated by equally elaborate, compound percussive events. These compound
sound events are made up of characteristic sounds used through the piece, for example a struck piano frame is combined with a plucked string which has glissando both up and down simultaneously, plus a segment of the high-frequency duplet.

Another scraping noise, at 11:32.31, seems to trigger some middle range chimes, and a C# pitch centricity is reinforced throughout. The development in the latter part of the section is hinted at when the high chimes, at 11:54.71, explore a three pitch motive in a way that is reminiscent of what we heard earlier in the section.

The final phrase in the section is initiated by another struck piano frame sound at 12:07.07. This unleashes a rich texture of a sustained Wurlitzer organ-like chord that adds to the C# centricity with further pitches and intensified amplitude modulation. After a slight relaxation halfway through the phrase, an even richer version of the same timbre erupts by exploring the three pitch motive idea, employing varied amplitude modulation, and exaggerating pitch and variance as the vibrato rate increases over its duration. Finally the bowed metal-bowed glass dialog returns at the end of the section to be elaborated on in the final section.

Figure 6.8: Wind Chimes study score 12’-14’
6.5.9 Section 9

[12:50.00-15:05.00]

**Observations**

This final section begins with a low, sustained gong-like sound which rises and falls in amplitude for most of the final section up until 14:30.00. Its pitch seems to vary from time to time between 238 Hz (A#3+39) and 197.4 Hz (G3+12).

Beginning at 12:52.73, there is almost a literal repetition of the bowed metal-bowed glass interplay of sounds we heard at the end of Section 8. In Section 9, however, we hear quite an elaboration of the interplay, with some transposition in pitch. In its first incantation, the pitch is centred on 1039.3 Hz (C6-11).

At 13:06.10, we hear a first variation of the bowed metal-bowed glass interplay transposed to 868.3 Hz (A5-23), followed by a shortened motive at 1039.3 Hz (C6-11). What follows at 13:20.00 until 13:29.20 is a repetition of the 868.3 Hz (A5-23) figure, and then the same figure elongated and varied from 13:29.20 until 13:48.95. Another significant frequency at this time is 752 Hz (F#5+28), and being somewhat over a whole tone lower than the (A5-23), there arises some resultant beating.

At 13:48.95, the shortened (C6-11) motive is heard again, followed by quite an elaborate passage exploring variations on the (A5-23) bowed metal-bowed glass interplay. This continues for over 35 seconds and, as it is just dieing out, it is interrupted by a loud, low, plucked string-like sound at 14:30.88. It has significant frequencies of 170.6 Hz (F3-40), 203.55 Hz (G#3-34), and 229.13 Hz (A#3-29).

The gong sound is no longer present but there is a very low-frequency rumble that continues until the end of the piece. Shortly after the plucked string sound, we hear the re-emergence of a complex organ-like high-frequency triplet combination at 14:32.70, lasting until 14:50.00. As it begins to fade we hear some softer repeats of the bowed metal-bowed glass sounds fading as the low-frequency rumble fades and the piece comes to an end.
Discussion

This final section really acts as a coda to the work. The interesting timbral interplay between the bowed metal and bowed glass sounds encountered in the previous section is elaborated much further in the final section. Here we see “variations on a timbral theme” with transposition by a minor third added for further variation.

The low gong sound acts like a timbral cantus firmus, underpinning the higher sounds throughout. Smalley is once again utilising the entire spectrum. The re-introduction of a plucked string acts as a point of punctuation and introduces an unexpected elaboration near the finish. It announces the organ-like flourish and then the piece quietly winds down.

Figure 6.9: Wind Chimes study score 14'-End

6.6 Analysis of the Whole Work

So far we have analysed the work from moment to moment in a serial fashion – section by section. According to our SIAM model, we have segregated the various sound events for each section (mostly documented in the “observation” passages above), and we have looked for horizontal and vertical instances of integration for
each section (noted in the “discussion” passages above). Now we must review the whole work and try to detect and describe global structures and recurrent syntactic forces operating within the work.

We are at the point where we want to make use of the “assimilation” and “meaning” components of the SIAM framework. That part of the framework says that we need to consider: the nature and type of discourse, global organization in time, hierarchical relationships, and implication-realization.

At this stage we could say that there are at least two possible views of the work, as prescribed by the SIAM model, and they are a hierarchical view and an implication-realisation view. A hierarchical view suggests that the work can be interpreted at the short-term (temporal) level as a relationship of isolated events or concepts, at a medium-term (temporal) level as a relationship between phrases or intermediate concepts, and at the long-term (temporal) level as a relationship between broad sections or high level concepts. An implication-realization view implies that the work is interpreted in a more dynamic way from moment to moment. But equally we could assert that it could involve a more dynamic interpretation of the hierarchical model where the focus shifts frequently and almost simultaneously through the hierarchical levels. It may involve keeping track of various phenomena at different time scales (short-term, medium-term, and long-term), and interpreting differing implications at the level of those time scales. Implication involves expectation, and expectation requires memory. So we will concentrate on the discovery of relationships at different time quanta, and the methods required.

How can we discover hierarchical relationships? We can “reduce” the time-scale by distilling the full score into a “short score”, and thereby create an additional meta-level in the analysis - a hierarchy of interpretation. At this higher level of interpretation we can make assessments of whether semantic, syntactic, or acoustic factors are the basis for hierarchical formation, what other categorical generalisations seem to emerge, and how these various factors progress through time as the work unfolds.
By plotting how the categorical factors change over time, we can form a picture of the overall structure of the work. This approach is not dissimilar to the formulation of a Schenkarian analysis in Western tonal music.

The following section describes the methodology employed to create a hierarchical interpretation through the creation of a “short score”.

6.6.1 Methodology

Having created an interactive study score, a hard copy of each two-minute screen was printed and they were joined together. Sections were marked on the printed study score. Annotations from the observations and discussion passages were added for each section. Having segregated the sound events through representation in the study score, this step was meant to highlight the functional relationships between the sound events (entities), and thereby concentrated on the integration aspects. The result of this step was an annotated study score.

The next step was to reduce an eight-page annotated study score to a four-page analytical reduction or short score. Remember that the idea of this step was to distil the full score into a shorter score, so that we could create another meta-level in the analysis in order to create a hierarchy in the interpretation (time-span reduction).

Reduction to short score involved:

- The recording of the major features from each section of the work.
- The grouping of those major features into categories.
- An examination of the progression of major features from section to section.

As the reduction technique proceeded, the list of major features seemed to group into the following categories:

- Texture
- Frequency related organisation
- Time related organisation
• Special features

• Semantic reference versus a syntactic reference

A table was created to plot the way these categories change from section to section. A timeline was included. Sections 1 and 2 were collapsed together to form just the one section of duration 1:22.87. The remaining section segmentations seemed appropriate.

The net result of the reduction process was to create a four page short score from the eight screens of the interactive study score.

![Figure 6.10: Tabulation of categorical features for sections of Wind Chimes – Page 1: 0’ – 4’](image)

![Figure 6.11: Tabulation of categorical features for sections of Wind Chimes – Page 2: 4’ – 8’](image)
6.6.2 Progression within the categories

Texture

Right from the very beginning there is a strong use of the attack-continuant model that is subsequently embellished and transformed.

From the tables we can observe that, from Section 3, there is an increasing use of terms such as “spectrally pitched”, “spectrally dense and rhythmic”, “timbral counterpoint”. The use of these terms is meant to illustrate a connection to traditional Western art music syntax that seems to be influential in the organization of frequency and time.
This was confirmed during a conversation with Smalley (2004b):

There always seems to be some kind of pitch relationship or potential within the material. It may be realised or not realised, and particularly less now in more recent pieces. I’m working with intervals and centres, with potentially tonal allusions, but relating those to harmonicity, using the word more loosely, in a timbral approach. I think all my pieces, to a greater or lesser extent, have some kind of slipping in and out of a pitch framework, but I wouldn’t say that it is pre-organised. It’s a sort of instinctive thing, and I think it is probably compositional habits.

I slip into ways of working in method or approach that automatically harness that kind of thing. For example in *Wind Chimes*, I wouldn’t be able to tell you necessarily whether pitches that may be central later in the piece relate to pitches that are earlier in the piece.

I may have arrived there by some logical sequence of events - in the short to middle term temporal scale. I haven’t necessarily designed a structure which has an underlying scheme. So, yes, you’re right, there are note centres or pitch centres. The note itself may or may not matter, depending on whether it has a harmonic relationship to the surroundings. At one stage in *Wind Chimes*, there is a ninth chord that has various motions that obviously is a harmonic moment or gives a harmonic message. I can’t remember how I get there or leave it. It might be done in a logical sort of way or it might be disrupted. The other thing about that kind of situation is that it allows some kind of instability, but is arrived at as a part of the structure, as a part of the system that remains stable for a time and allows you to follow other things. This notion of stability allows you to play with other elements. The other elements allow the ear to penetrate and listen to the other elements although pitch might be providing the underlying foundation. So it enables you to play segments of variations where the pitch remains stable and you’re varying the musical structure other than pitch – they might be spectral variations. There are covertly or overtly tonal episodes or tonal arrivals moving in or out. By tonal I mean varying more than one pitch and where there’s an interval there may be a harmonic implication or that the interval might itself have a certain kind of association.

In Smalley’s case, the traditional Western art music relationships have been pushed higher in the spectrum, so that the relationships are between higher partials, and hence the use of the terms “timbral” and “spectral” in conjunction with counterpoint and harmony. The musical operations also seem disguised by the cloak and mask of semantic reference.

As the work progresses further, we can observe that, beginning around Section 6, the initial attack-continuant framework becomes transformed into more of a compression-relaxation style. The tension is increasing within the work as sections, and phrases
within sections, overlap with each other. There is any increased density of activity as the work progresses. Tension may be caused through the interruption of expectation, where the expectation is of a regular statement in time. The regularity is truncated and overlapped by other events – causing the interruption. Smalley (2004b) describes his process this way:

As you would realize from my articles, I am setting up expectations and implications, or I hope I am with the language that I am trying to use, which is the gesture-texture time relationships of the kinds of events that you're looking at – to harness those in such a way to create a “narrative” approach to form.

The activity and increased use of processing build to a climax in Sections 7 and 8. Then there is a winding down in Section 9 (timbral theme over a timbral cantus firmus). The climax of activity is very classical in its shape and form.

**Frequency Organisation**

Already we have coined the term pitch centricity to depict the gravitational point of attraction that characterises the pitch of sound events within wind chimes. The use of pitch transposition by Smalley is not a blatant use of tonality with sound samples as source material, but rather it is a subtle phenomenon that is used functionally to provide connection and continuity between seemingly separate sound events.

Smalley uses the Gestalt principle of proximity in frequency as a bonding material between the sounds. By superimposing a certain fluidity in pitch on the primary source materials, the resultant “frequency bonding” between them breaks down the “source bonding” that the individual sounds may exhibit when heard in isolation.

Beginning with a pitch centricity of (A#+50), there is a move to B natural which becomes the prime pitch centric force around which the remainder of the piece is structured. At the centre of Section 3 we have a pseudo cadence of G-E, then a return to B, and an elaboration to E and back to B (I-IV-I?), and then a spectral chorale at the end of Section 5 (ca 6:48.00).

Section 6 begins on B with an excursion to its neighbour of C#, back again, and then a rise of a perfect fourth to E.
Section 7 meanders around E/D#/D/#E, while Section 8 plays with some diminished triads and a predominantly C# pitch centricity. There is the possibility that one could interpret Sections 7 and 8 as a stepwise dissent from E-D#-D-C#, with a final descent of a third from C# to A#, and then the C to A sequence.

The piece finally gravitates back to the regional pitch centre, or a slightly flattened version of the original pitch centre.

As noted above, Smalley uses the principle of proximity in frequency as a binding technique. In traditional harmony this could be expressed as using sustained tones or stepwise motion between adjacent chords. Where there is a leap in frequency (or pitch), Smalley users the consonant intervals of a third (and sixth) or a perfect fourth (and perfect fifth), for example G-E, or B-E.

Both traditional harmony principles and Smalley's syntax may be tapping into acoustic properties in the application of these techniques.

**Organisation in Time**

Wind Chimes was specifically chosen because it does not possess a dominant beat or pulse-related rhythmic organisation. However the work does contain some elaborate rhythmic features. On a macro level, organisation in time is achieved through a sectional structure that starts to blur and overlap with itself in the second half of the work. On the micro level, there are distinctive rhythmic motives that appear early in the piece, and become a recurring signature throughout, being developed and transformed in the manner of a traditional development movement, or perhaps theme and variations.

Smalley reserves the high-frequency portion of the spectrum for these rhythmic motives, further demonstrating how he takes traditional musical devices and pushes them into the upper realms of the frequency spectrum.

What we have identified as the high-frequency triplets and duplets are used again and again throughout the work. Smalley has since confirmed their source is a four note rhythmic figure, where the final sound is softer than the preceding three.
Other time related techniques come into play, from Section 5 onwards, with increasing use and fragmentation. Repetition and variation is used liberally throughout (eg. The accelerating “choofs”).

There is pre-echo or the premonition of an anticipated sound, quieter ghost components of previous sounds, repetition of sounds in a new context, and the return of strategic sounds at critical points, for example the return of the opening sound on its own at 6:42.06. These strategic or distinctive sounds provide a scaffold for other sounds to be positioned around.

In Section 6, short-term rhythmic elements like the accelerating choofs and bell sounds are mixed over very long sustained sounds to heighten the compression-relaxation schema before an explosion of sound at 7:09.22.

Compound sounds are used throughout the work displaying the Gestalt effects of proximity in time along with the triggering effects created by events that are nearly coincident. Here proximity in time influences some sort of causal interpretation between sounds by the listener, creating the effect of a gesture.

Compound attack-continuant sounds are heard in Section 7 (beginning at 7:51.35), and an elaborate exploration of call-response rhythmic interplay between what we’ve called bowed metal and bowed glass sounds arises in Section 7 and continues throughout the rest of the work.

Loud percussive sounds like foot-stomps, struck piano frames, bass drums and so on are used to mark sections and provide punctuation points. They are also used to initiate phrases and scraping noises are used to trigger continuous sounds. In Section 9 (beginning at 12:50.00), the density of sound events thins out as rhythmic elements like the bowed metal bowed glass interplay wind down. These are punctuated with distinctive iconic sounds such as the plucked string and the organ-like flourish.

**Semantic and Syntactic Progression**

The final categorical grouping considers both the semantic aspects of the work and the syntactic factors operating throughout. Our earlier model of the network of relations involved in the interpretation of acousmatic music placed a semantic network at one
end of a continuous scale, and syntactic network at the other extreme (see Figure 5.1). Our semantic network is stimulated when a possible source or cause is associated with a particular sound event. Sections of music with semantic associations are said to be source-cause discourse dominant. In this case, relationships between sound events will tend to be semantic relationships, although there can be some overlap between semantic factors, ecological relationships, physical referents and the acoustic factors contributing to the recognition of certain sound events.

Semantic relationships are associated with a real environmental space - the concrete. Semantic relationships are characterised by the operation of association, which is termed indexical representation according to some semiotic models.

The other end of the scale is the syntactic network where a typological-relational discourse is dominant, although semantic reference may be present. Relationships are established through some operative syntax superimposed on sonic materials. Because we are dealing with the organisation of sound, there must be some relationships between the syntax used and the temporal and spectral elements (acoustic factors), although the amount or degree of coupling can vary. This abstract end of the scale is associated with a symbolic semiotic interpretation.

At the beginning of Wind Chimes, we are presented with several semantic “characters”. The opening chime sound, the low choofs, the gong sound and the struck piano frame all reappear at regular points throughout the ensuing “drama”, although usually in an altered form.

The piece begins with these, mostly untreated, separate sounds. The form of the work is that the sounds become more processed as the work proceeds. There is continual reuse of early material, but with heavier signal processing and denser use of sound.

The effect of more signal processing is to sever the source-cause bonding. When this bonding is severed, some other type of bonding must be substituted for it, otherwise the work would become a collection of unrelated events. We can characterise some of Smalley’s syntax as a modified form of traditional Western art music syntax. Tonality is replaced by pitch centricity in Smalley’s syntax. Sounds are transposed to similar pitch centres, we can call these centres of gravity perhaps, then they are bonded
together by virtue of their frequency proximity and functional harmonic proximity. Here we see a reinforcement of the musical syntax by acoustic properties and perceptual factors. There is a happy confluence of systems working together.

It is no accident that the use of a maximum transformation of concrete materials is coincident with the climax of the work. The form of the piece is that it begins with concrete elements which are readily discernible, and progressively moving to an abstract musical space where the spectral and temporal relationships dominate. Then the work relaxes to where individual concrete sounds become recognisable once again near the end.

When the distinctive concrete sounds used earlier in the work reappear in an untreated form at periodic junctures throughout the work, they are generally used as structural signposts at strategic or critical moments - for example the original chime sound returns solo at 6:42.06. One could view this as a golden section time point or as a counterbalance to the climax at around 9:40.00.

Smalley (2004b) has confirmed himself that he is using a traditional musical narrative approach with periods of tension and release:

Yes, my music definitely operates within a traditional teleological framework of tension and relaxation, which of course is the situation within traditional Western art music, and any times I’ve tried to go outside it I haven’t managed to do it. For me that tension and relaxation organizational method is one of the affective records of Western art music and is fundamentally related to our psychological perception of time. I can’t personally do away with it. I wouldn’t adopt a continuously repetitive framework, well I could, but it would have to, in some way, consider the tension-relaxation paradigm. I couldn’t for example, present a slice of recorded environment. I couldn’t do that in a non-teleological way. I would hear it or use it in some way that appeals to our musical senses at the psychological level.

This tension-release pattern is mirrored in his heavy use of the attack-continuant model on the micro scale, and the compression-relaxation accumulation of percussive short sounds followed by layers of long sustained sounds on the macro-time scale.

Forward motion is provided by connection between successive sounds in the frequency domain coupled with an increasing density of sound event activity, and
reinforced with the increasing use of signal processing. The micro schema reinforce the macro formal organisation.

6.7 Summary

The interpretation of the analytical results has resulted in the emergence of certain syntactic traits: compound sound events; stratification of the frequency spectrum; pitch centricity and harmonic fields; organisation in time; semantic and syntactic progression.

6.7.1 Compound sound events

Smalley uses compound sound events where the technique may consist of:

1. The attack portion of a sound event (e.g. 4:49.54) contains a number of components drawn from different sound sources.

2. The sustain portion of a compound sound event draws out selected frequency components, from the attack, that contribute to a certain "pitch centricity". Other sound entities contribute to the "compound sustain" to add colour, variety and complexity.

3. A variation of the "compound attack" is the "compound gesture" (e.g. around 4:58.80). Sound elements are accumulated around the one point in time, often creating a complex gesture that may have no counterpart in the real world. The "compound gesture" becomes a new creation in Smalley's musical world.

6.7.2 Stratification of the frequency spectrum

The use of many compound events can use a large proportion of the frequency spectrum, so Smalley avoids a crowded, muddied sound, and the ensuing effects of masking, through a skilful selection of contributing sounds where each one occupies its own segment of the spectrum in a way that doesn't overlap too much with others.

In general, Smalley uses the whole spectrum, with sustained sounds in the lower register, pitched sounds in the middle register, and interesting noise colourations in the high frequency register. Of course he uses combinations as well.
6.7.3 Pitch centricity and harmonic fields

Smalley has moved the manipulation of pitch upwards within the spectra of sounds used. Rather than creating melodies that consist of notes with definite pitches, Smalley emphasises certain frequencies within the spectra of the sounds he employs, or he transposes sounds so that they gravitate to certain pitch centres.

These pitch centres act like pedal points in tonal instrumental music. At around 05:22.00, one such pitch centre is a quarter tone above E natural. Smalley creates a "harmonic field" using the pitches associated with an E major/minor seventh chord.

A B-natural pitch centre is alternated with the E-natural pitch centre throughout the section. Smalley colours these with thirds, fifths, and sevenths to create his harmonic fields.

Smalley has also re-interpreted traditional harmony and voice-leading practices by taking the principle of "common tones" between two successive chords in tonal harmony and applying it as a principle of "common partials" between two successive sound events in acousmatic music. The common partials then become a binding agent between the sounds through a strong horizontal streaming tendency.

An extension of this principle is the principle of "stepwise motion" used in harmonic progression. Smalley creates a stepwise motion of sustained partials when some dominant partials are changed to values that are close in frequency.

A beautiful application of these principles is evident at the end of “Section 5” (at 06:23.04) where we have a "spectral chorale". The B natural tonality is embellished by a G harmony with motion to B’s neighbour C# and back again.

Just like the relationship between speech and song, Smalley is giving voice to ordinary sounds by bringing out the voiced components in those sounds. He achieves this by either reinforcing certain frequencies, transposing to a particular pitch, or super-imposing selected sustained frequencies.
6.7.4 Organisation in time

Throughout Wind Chimes there is no strong pulse, no meter, and no elaborate rhythmic structuring present, however there is careful structuring in time. There are a number of temporal elements that are apparent.

The attack-continuant paradigm is a strong structuring force. Pseudo-rhythmic figures are used within the attack-continuant framework. Some examples are the high frequency couplets and triplets beginning at 4:58.80, or the low frequency figure at 5:05.39. Sustained partials are overlapped with the next attack-continuant expansion, which provides a "proximity in time" Gestalt binding force.

Simultaneous attacks create "compound sounds" by virtue of their "vertical integration". Nearly-simultaneous attacks create "compound gestures" that are so unique they generate interest and variety.

Repetition is also used in a number of ways. Repeating a sound at short, and ever-decreasing, time intervals produces a feeling of acceleration, propelling the listener towards a goal event. A more subtle technique, used over a wider time-frame, is the creation of a "pre-echo" or anticipation of a sound. Smalley achieves this by taking its spectrum, transforming it in some way, then placing the processed sound before the unprocessed sound in time. For example the "ghost components" at 5:21.59 are from the high cymbal at 6:04.47. This is much more subtle and effective that using the reverse order, that is playing the original sound followed by its transformations. Repetition of sound in a new context is also extensively used, for example the "whispering glass rim" sounds at 5:23.80 are repeated at 5:48.36.

Distinctive sounds make judicious return appearances throughout the piece. For example, the return of the original percussive chime from the start of the piece returns at time 6:07.17. This is disguised with a busy phrase, but it is like a preparation of a full, unaccompanied return in the next section. The use of these distinctive sounds has been given the term “scaffolding”. The distinctive sounds provide a skeletal framework that is fleshed out through processing and progression. Loud percussive sounds (like foot-stomps, struck piano frames, and bass drums) are used to provide punctuation points and mark sections.
6.7.5 Semantic and syntactic progression

The form of the piece is that it begins with concrete elements which are readily
discernible, and progressively moves to an abstract musical space where the spectral
and temporal relationships dominate. This is achieved through an increasing use of
signal processing, transposition, and an increase in density of sounds as the work
progresses. There is a climax in the work and then it relaxes to where individual
cement concrete sounds become recognisable once again near the end.

When the distinctive concrete sounds, used earlier in the work, reappear in an
untreated form at periodic junctures throughout the work, they are generally used as
structural signposts. We have used the term “scaffolding” to describe this
phenomenon.

Smalley uses a classical Western musical narrative approach with periods of tension
and release. On the micro scale, the tension-release pattern is mirrored in the
predominant use of the attack-continuant model. On the macro time scale, the
compression-relaxation accumulation of percussive short sounds followed by layers of
long sustained sounds represents another variation of the tension-release pattern.

Forward motion is provided by connection between successive sounds via proximity
in the frequency domain coupled with an increasing density of sound event activity
which builds excitement. These tendencies are reinforced through the increasing use
of signal processing. We can see that schema employed at the micro level reinforce
the formal organisation at the macro level.

6.8 Conclusions

Reflecting on the process of analysis, one can observe that the application of the
SIAM framework is not a systematic, stepwise, serial sequence of actions, but rather a
checklist that is accessed according to what may be appropriate for a particular
segment of the work.

In the segmentation and sound event identification process, the very first impulse was
to ascribe a label to a sound event, either as an identification of the possible cause or
as an onomatopoeic reference to the sound event. This reflects how strong the urge for
us to identify a possible source-cause for a sound is, and it provides a convenient
device to use when we are recording and communicating our observations and
thoughts in text.

What is immediately obvious is that Smalley is playing with the internal spectral
structures of what we think of as single sound events. The listener is drawn into a
giant internal sound world within which Smalley performs his microsurgery. One
minute Smalley is a surgeon and excising components from the spectrum, the next
minute he is a sculptor extruding and elongating individual frequency components.

Smalley has taken some traditional tonal music concepts and pushed them higher in
the frequency spectrum, or in some cases embedded more in the spectrum, but lower
down. So we see harmony being expressed as a "spectral chorale", rhythm expressed
in the "HF Triplet" figures, and melody in the form of filtering of low frequency
material to create ascending and descending melodic lines, or "spectral counterpoint".
He uses the notion of compound sound events extensively where several simple
sounds add together to give a complex composite.

*Wind Chimes* also makes extensive use of attack-resonance, which may involve a
combination of several of the above techniques. An attack that is an agglomeration of
sounds is prolonged into a resonance phase by the extension or addition of some
component frequencies. At times, the form of the work is like an attack-resonance on
a macro scale. Sound events tend to aggregate at certain time points then there is a
relaxation of activity revealing long sustained sounds that have their own fascinating
micro-colourations. The form is like a breathing animal: a sudden in-breath, followed
by a momentary held-breath, then a long, lingering exhalation. The pattern is also
reminiscent of a gust of wind exciting a set of chimes in ever-increasing bouts of
intensity.

Overall, the initial attack-continuant framework becomes transformed into more of a
compression-relaxation style. The tension increases within the work as sections, and
phrases within sections, overlap with each other. There is an increased density of
activity as the work progresses.
The term "pitch centricity" has been employed to try and convey an observation that, although Smalley doesn't use an extensive tonal music organisation of pitch materials, he does manipulate pitch to create certain points of attraction that pitch may gravitate to. Transpositions tend to be by thirds or sixths too, which is wider than the critical band. On the other hand, Smalley uses smaller intervals for simultaneous events, less than a major second, to create beats and to provide colourations.

The overall form of the work is that it seems to move from the more concrete to the abstract. It progresses along the scale from a Source-Cause Discourse that is "about chimes and the wind" to a Typological-Relational Discourse that explores frequency and time relationships. This is achieved through an increasing use of signal processing and manipulation as the work progresses. Processing slowly detaches the sound events from their real-world connotations and places them in the new musical sound space Smalley is creating. The sound space uses pitch centricity as points of attraction for sounds. Singular sound events are bound together with a glue that is primarily frequency. Where simultaneous sounds share a frequency component they tend to bind together. Where sequential sounds share a frequency component, or are very close together in frequency, they tend to bind together to form a stream. So frequency components are extruded from one sound, elongated, then overlapped with the same frequency within another sound to fuse the events together to create a new gesture. The other binding force is time. Simultaneous events fuse together to form compound sounds and nearly simultaneous events tend to form an association where one sound seems to trigger the other implying some type of causal connection.

The form of the work is sectional and episodic. Smalley makes use of a lot of repetition of sound events, but the repetition is varied. Either the sound events are a processed repeat of the originals, or a literal repetition is set within a new context of different sounds around them.

The interpretation of the analytical results has resulted in the emergence of certain syntactic traits and these have been summarized in Section 6.7.
Chapter 7 : Summary and Conclusion

7.1 Framework for Analysis

The three research questions that were posed at the outset of the investigation, pertaining to the development of a framework for the analysis of acousmatic music, can be summarised more succinctly by the following:

Can a framework for the analysis of acousmatic music be derived from cognition theories, research on the auditory perception of everyday environmental sounds, and studies into the perception of Western tonal music? If so, what are the framework’s attributes?

In Chapter 5, all the constituents of such a framework were brought together. The framework has the following main elements: Segregation, Integration, Assimilation and Meaning (SIAM). To assist with assimilation, meaning, and the discussion of discourse within the work, an approach to listener interpretation and attention was developed (see Figure 5.1).

7.1.1 The SIAM Framework

Here is a brief recap of the SIAM framework:

Segregation: identification of sound events and the factors responsible for identification.

Integration – Horizontal: identify sequential streams and patterns of sonic objects; determine causal linkages, relationships and possible syntaxes; consider organisation in time and the horizontal integration of pitch.

Integration – Vertical: consider vertical integration and segregation as a cause of timbre and texture variance; consider psychoacoustic dissonance and musical dissonance; consider emergent properties relating to pitch (horizontal overlap).

Assimilation and Meaning: consider discourse on the source-cause dominant (semantic) to typological-relational dominant (syntactic) continuum; consider
global organisation in time and any hierarchical relationships; consider expectation-interruption, arousal and meaning.

The broad form of this procedure is derived from the work of Bregman (1999) on auditory scene analysis, in combination with Bigand’s ideas on event structure processing (Bigand, 1993). The segregation and identification of sound objects also draws on the work of McAdams (1993). The establishment of factors responsible for sound segmentation and recognition has been influenced by the perceptual studies of everyday environmental sounds carried out by Gygi (2001), and by Howard and Ballas (1980). Identification of horizontal streams and the consideration of syntactical relationships between sound objects, and patterns of sound objects, has been informed by the experiments carried out by Ballas (1993). Considerations relating to vertical integration have been combined from Bregman (1999) and Smalley (1994). The notion of source-cause discourse derives from Smalley (1994), while the hierarchical “detection” methods of time-span reduction and prolongation reduction have been documented in Bigand (1993), and Lerdahl and Jackendoff (1983). Arousal and meaning is discussed in Dowling and Harwood (1986) and they refer to the symbolic interpretation ideas of Pierce (1931-35). A model of implication-realisation has been elaborated by Narmour (1989).

The listener interpretation-attention illustration (Figure 5.1) provides the context within which the musical work resides. It is an attempt to define a cultural space for acousmatic music that highlights the crucial role of the attitude of the listener. The study has arrived at a multi-layered depiction of possible listener interpretations, represented as a set of dichotomies. Each dichotomy has two poles separated by a continuum. One continuum explores the world of signs and symbols in that it borrows the terms defined by Pierce (1931-35) and elucidated by Dowling and Harwood (1986). The next continuum uses the kind of terminology developed within the tradition of musique concrète, inspired by Pierre Schaeffer and extended by the work of Smalley (1994). Whereas Schaeffer documented sound object “typology” extensively (static emphasis), Smalley has written a great deal on “morphology”, or the way sounds change with time and the consequential implications of their transformation (dynamic emphasis). This second continuum is from source-cause
discourse dominance to typological-relational dominance. Composers exploring expression along this continuum play with the “operation” of source bonding.

The next continuum, from “semantic network” to “syntactic network”, has been informed by the work of psychologists and it seeks to document the factors responsible for sound recognition. Starting from the left we recognise sounds from the world of our everyday experience, we can give a name to them and we can ascribe them a meaning. As we move along the continuum to the right, the ecological psychologists would argue that we may not recognise other sounds directly, but we can make inferences about the nature of the sounding material and the way the material has been excited. These are physical factors telling us something about the sounds and their environment. Further to the right, acoustic factors dominate and the spectral and temporal information can activate syntactic networks that establish temporal patterns between sounds in a sequence.

In the final continuum, there is an attempt to show that the concrete world of a real environmental space can operate in a very fluid-like relationship with the abstract realm of a purely fictional musical space.

Whether consciously or sub-consciously, a listener can move freely through all of these layers of interpretation while attending to an acousmatic work. Perhaps the interpretation continua could be identified as different listening modes. It should be noted that these modes are not exhaustive in that there exist a myriad of other listening modes that haven’t been highlighted here. What has been highlighted are the kinds of interpretation that may have a useful relationship with the “culture” of the acousmatic music genre, derived through the literature on acousmatic music and the psychology of the listening process.

To round off this section, the assertion can be made that:

A framework for the analysis of acousmatic music can be derived from cognition theories, research on the auditory perception of everyday environmental sounds, and studies into the perception of Western tonal music.
The attributes of the framework, derived in previous chapters, have been re-iterated here, along with the sources for their origins. The questions pertaining to the application of the framework can now be addressed.

### 7.2 Application of the Framework

In the introduction chapter, a second set of research questions were concerned with the practical application of a cognitive framework to the analysis of acousmatic music:

- What methods and tools must be adopted to apply the cognitive framework to the practical analysis of repertoire acousmatic works?

- What notational forms must be derived to communicate any analytical findings in a written medium?

- What new insights result from the application of the constructed framework and methodology?

- Is the framework an effective approach to the analysis of acousmatic music?

- What are the implications for subsequent research?

Each one of these questions will now be addressed.

#### 7.2.1 Methods and Tools

The first component of the defined methodology was to carefully consider the aims of the study, which were to apply the SIAM framework in an analysis of the work *Wind Chimes* in order to: Discover insights into the work itself, discover the syntactic forces that may operate between sound events within the work, and to test the appropriateness and usability of the SIAM framework.

The analytical method was a combination of signal analysis and critical listening. A number of different types of signal analysis and representation methods were trialled before a sonogram representation was settled on. It was deemed to be the most useful in terms of sound event segmentation and frequency representation. The program Adobe Audition was used to play the sound file and to create various spectral displays of frequency versus time at different resolutions. To enhance the listening process, the
graphical programming language PD (Puckette, 1996) was used to assist with “tuning in” to significant frequencies and pitches aurally.

The work was segmented into sections taking into account several factors including human long-term memory considerations, the "sliding window" notion of our short-term working memory, usable screen size displays, and factors associated with the musical work itself, such as segmentation at obvious sonic boundaries like moments of silence or passages with long sustained sounds. These boundaries defined a series of nine “sections” within the work, which were numbered for easy reference.

Analysis proceeded in a linear way from start to finish. For each section, a set of observations was recorded in text form, and a pictorial representation of separated sound events was drawn in pencil on paper along a timeline. Analytical data was also drawn on the pictorial representation. The data that was collected included an event's: start time; duration; perceived pitch or significant frequency components or both; graphical indications of amplitude envelope; graphic symbols depicting special features, e.g. pitch glissando. A "discussion" passage was also written for each section in order to interpret the observations that were made.

Once the observations, pencilled pictorial representation, and discussion passages had been completed, the question of how to record and communicate observations then arose, and so a Flash-based "interactive study score" was developed to provide a dynamic visual representation along with synchronized playback of the recorded work being analysed.

Having analysed the work from moment to moment in a serial fashion according to the SIAM framework, the whole work was reviewed in order to observe and describe global structures and recurrent syntactic forces operating within the work. That part of the framework says that the following factors need to be considered: the nature and type of discourse, global organization in time, hierarchical relationships, and implication-realization.

In order to discover hierarchical relationships, the time-scale was “reduced” by distilling the full study score into a “short score”. This step was designed to facilitate assessments concerning semantic factors, syntactic factors, acoustic factors,
hierarchical formation, categorical generalisations, including how these various factors progress through time, and how they relate to each other.

In practical terms, a hard copy of each two-minute screen was printed, joined together, and annotations from the discussion passages were added for each section. Having segregated the sound events through representation in the study score, the discussion annotations were meant to highlight the functional relationships between the sound events, and thereby concentrate on the integration aspects. The result of this process was an annotated study score.

The next step was to reduce an eight-page annotated study score to a four-page analytical reduction or short score.

Reduction to short score involved:

• The recording of the major features from each section of the work.

• The grouping of those major features into categories.

• Examining the progression of major features from section to section.

As the reduction technique proceeded, the list of major features seemed to group into the following categories: texture; frequency related organisation; time related organisation; special features; semantic reference versus a syntactic reference.

A table was created to plot the way these categories changed from section to section, and the net result of the reduction process was a four page short score created from the eight screens of the interactive study score. Conclusions were then drawn regarding macro structures operating within the work.

Reflecting on the process of analysis, one can observe that the application of the framework is not a stepwise, serial sequence of actions, but rather a checklist that is accessed according to what may be appropriate for a particular segment of the work.

Most of the analytical process was carried out by hand, and was very time-consuming. Segmenting the sound events, recording times and durations, trying to determine salient frequencies in the midst of complex textures, and then transcribing the results,
could take a full day’s work just to analyse 30 seconds of the piece. This very labour-intensive approach could be facilitated by some sort of data reduction techniques, such as auditory modelling, or a semi-automated approach, for example the use of some auto-correlation methods to facilitate salient pitch detection (see below).

Concentrated, critical listening was the primary means of analytical investigation, but the signal analysis techniques greatly enhanced both the detection of phenomena and the specification of hard data. One of the under-stated aims of the investigation was to provide enough of this hard data, in terms of exact timings and frequencies, so that another researcher could view the study score and create their own interpretation of the work.

Refinements to the means of representation could be made, but this will be taken up in the next section.

7.2.2 Notational Forms

The question addressed in this section is: What notational forms must be derived to communicate any analytical findings in a written medium?

In deriving notational forms, consideration has been given to the communication of analytical results, and to the use of notation to assist in the analysis itself.

A number of different signal analysis and representation methods were trialled. They included spectral analysis, fundamental frequency analysis (including auto-correlation), amplitude envelope trace, and noise analysis. Of all of these forms, the spectrogram (or sonogram) was found to be the most useful in terms of sound event segregation. High resolution spectrograms were created and a combination of visual inspection and aural tuning (using the oscillator bank patch) assisted in the identification of relevant frequency components.

To communicate the findings of this segregation-integration process, several types of representations were devised:

- An interactive study score.
- An annotated study score.
Interactive Study Score

The interactive study score divided the work up into two minute segments for representation on a visual display screen. With time displayed horizontally, the screen was divided vertically into a sonogram display area at the top, and an annotated graphic symbol area below (see Figure 6.1). The sonogram area was further divided into a full spectrum display above and a part-spectrum display underneath. The zoomed-in version of the spectrum, below 1500 Hertz, was useful for concentrating on pitch aspects.

The area underneath the sonogram is meant to depict the segmentation of the work into separate sound events, and to provide some hard data on those sound events, such as important frequencies and pitches, as well as descriptive information (e.g. “gong-like”). A grid background was constructed to facilitate alignment with the time scale of the sonogram plots. Horizontal placement of graphic symbols represents fairly accurate representation of an event’s start time and approximate duration.

The playback controls are used to initiate the playback of the sound track of the piece in synchrony with the visual representation. The ability to start and stop playback at will facilitates the isolation of specific events at particular time points within the whole work. In a sense the study score is depicting the work “out of real-time” in order to enhance analytical investigation. It is mainly descriptive of events and doesn’t necessarily show the relationships between events. This issue is taken up in the next section on the “annotated study score”.

Annotated Study Score

A hard copy “annotated study score” was created from screen grabs of the interactive study score. The purpose of this step was to highlight the functional relationships between the sound events and thereby concentrate on the integration aspects. Sections were marked on the printed study score, and annotations from the observations and discussion passages were added for each section. Lines and other markings linking related events were also added. One thing that should be noted here is that annotation of relationships for representation on the submitted study score interactive was
considered, but the amount of information was deemed to result in visual complexity to the point where it would cause visual confusion, rather than elegant illustration.

The annotated study score can be thought of as a working document that is a preparation for, and a transition to, the creation of a “short score” through the process of time-span reduction.

**Short Score**

The next step was to reduce an eight-page annotated study score to a four-page analytical reduction or “short score”. Remember that the idea of this step was to distil the full score into a shorter score, so that we could produce another meta-level in the analysis in order to create a hierarchy in the interpretation (time-span reduction), and begin to look at the progress of the whole work.

Reduction to short score involved:

- The recording of the major features from each section of the work.
- The grouping of those major features into categories.
- An examination of the progression of major features from section to section.

The list of major features grouped into the following categories:

- Texture
- Frequency related organisation
- Time related organisation
- Special features
- Semantic reference versus a syntactic reference

A table was created to plot the way these categories change from section to section. The net result of the reduction process was to create a four page short score from the eight screens of the interactive study score (see Figures 6.10-6.13).
The categories that emerged from grouping the major features of the work contained both confirmation of findings of previous researchers and some surprises. The emergence of two categories, one related to frequency organisation and one concerning organisation in time, was really no surprise and probably is a reflection of notions first espoused by Vanderveer (1979) that frequency information tells us something about the nature of the resonating body and time-related information tells us something about the nature of its excitation. In acousmatic music, frequency relationships create integration (or contrast) and time relationships create patterning.

The category labelled with the term “texture” could have been equally well described as “structural features”. In this category we find terms like “attack-continuant model” and “spectrally dense and rhythmic” and “scaffolding and triggering sounds used over sustained sounds”. These are all terms describing the way different sections are structured and perhaps the way different sonic elements are “layered”. Indeed the notion of fore-ground, middle-ground and background hasn’t been highlighted since this may be too simplistic for acousmatic where integration is a fundamental structural technique. An analogy to traditional contrapuntal textures seems more appropriate than to homophonic textures. Some writers have chosen to elaborate further on the fore-ground/background idea (Van Leeuwen, 1999).

The category of “semantic vs syntactic” features is an obvious but important one for acousmatic music. The way these forces compete or reinforce each other through the course of the work is critical to this genre of music. Their interpretation is provided in the “insights” section below.

Finally the “special features” category is a catch-all for a lot of the unique, but important, aspects that cannot be allocated to the above categories. For Smalley, the use of compound sounds and the re-use of material are a couple of special features that stand out.

In general, one can conclude that the interactive study score is extremely useful in the analytical process in “freezing” the work in time and helping to describe, in a visual form, the segregated sound events that have been discovered. The combination of spectrogram display, graphical symbols, textual and numerical data is also a useful
combination of data in the communication of event information, as is the inclusion of two resolutions of spectrogram data.

A better format to trial in the future may be to have full-screen spectrogram representations and superimpose symbols and data actually on the spectrogram. While it could be argued that this may erode the depiction of integrated single events, since some events may be spread across the whole spectrum while others may occupy a small segment of it, annotating the spectrogram could assist with the representation of relationships between frequencies and time information. Colour coding could also be useful.

The tabulation of a short score, through time-span reduction, assisted with the identification and communication of trends across larger portions of the work. The development of an annotated study score certainly facilitated the creation of the short score.

7.2.3 Insights

What new insights result from the application of the constructed framework and methodology?

Summarised here are the main conclusions regarding the insights gained into the organisation of Denis Smalley’s *Wind Chimes*:

- In *Wind Chimes*, Smalley is playing with the internal spectral structures of what we normally think of as single sound events. The listener is drawn into an expansive internal sound world within which Smalley performs his microsurgery.

- Smalley has taken some traditional tonal music concepts and pushed them higher in the frequency spectrum, or in some cases embedded them deeper in the spectrum.

- Smalley uses the notion of compound sound events extensively where several simple sounds add together to produce a complex composite.
- *Wind Chimes* makes extensive use of the attack-resonance model. An attack, which could be an agglomeration of sounds, can be extended into a resonance phase through the prolongation, or addition, of selected component frequencies.

- At times, the form of the work is like an expression of the attack-resonance model on a macro scale. Sound events tend to aggregate at certain time points, then there is a relaxation of activity revealing long sustained sounds that have their own fascinating micro-colourations.

- Over time, the initial attack-continuant framework becomes transformed into more of a compression-relaxation style. The tension increases within the work as sections, and phrases within sections, begin to overlap with each other. There is any increased density of activity as the work progresses.

- The term "pitch centricity" has been coined to try and convey an observation that, although Smalley doesn't use an extensive tonal music organisation of pitch materials, he does manipulate pitch to create certain points of gravitational pull that frequencies may be attracted to.

- Transpositions tend to be by thirds or sixths - wider than the critical band. On the other hand, Smalley uses smaller intervals between simultaneous sounds, less than a major second, to create beats and to provide colouration.

- The overall form of the work is that it moves from the concrete to the abstract. It progresses along the scale from a Source-Cause Discourse that is "about chimes" to a Typological-Relational Discourse that explores frequency and time relationships. This is achieved through an increasing use of signal processing and sonic manipulation as the work progresses.

- The primary strong binding force is frequency. Where simultaneous sounds share a frequency component they tend to fuse together. Where sequential sounds share a frequency component, or are very close together in frequency, they tend to link together to form a stream. Frequency components are
extruded from one sound, elongated, then overlapped with the same frequency within another sound to fuse the events together to create a new gesture.

- The second strong binding force is time. Simultaneous events fuse together to form compound sounds, and nearly simultaneous events combine to form an association where one sound seems to trigger the other implying some common cause, or a cause-effect relationship.

- The form of the work is sectional and episodic. Smalley makes use of a lot of repetition of sound events, but the repetition is varied. The repeated sound events are either a processed version of the originals, or a literal repetition set within a new context of different surrounding sounds.

- Finally, the interpretation of the analytical data has resulted in the emergence of certain syntactic traits: compound sound events; stratification of the frequency spectrum; pitch centricity and harmonic fields; organisation in time; semantic and syntactic progression. Each one of these phenomena will be briefly recapitulated below.

**Compound sound events**

Smalley uses compound sound events where the technique may consist of a “compound attack” containing a number of different sound sources, or a "compound sustain" to add colour, variety and complexity, or a "compound gesture" where sound elements are accumulated around a single point in time, but are not exactly coincident.

**Stratification of the frequency spectrum**

The use of many compound events can use a large proportion of the frequency spectrum, so how does Smalley avoid a crowded, muddied sound, and the ensuing effects of masking? He does this through a skilful selection of contributing sounds where each one occupies its own segment of the spectrum in a way that doesn't overlap too much with others. In general, Smalley uses the whole spectrum, with sustained sounds in the lower register, pitched sounds in the middle register, and interesting noise colourations in the high frequency register.
**Pitch centricity and harmonic fields**

In some instances, Smalley has moved the manipulation of pitch higher within the spectrum of a sound. Rather than creating melodies that consist of notes with definite pitches from a scale, Smalley emphasises certain frequencies within the spectra of the sounds he employs, or he transposes sounds so that they gravitate to certain pitch centres.

Smalley has also re-interpreted traditional harmony and voice-leading practices by taking the principle of "common tones" between two successive chords in tonal harmony, and applying it as a principle of "common partials" between two successive sound events within the acousmatic musical texture. The common partials then become a binding agent between the sounds through a strong horizontal streaming tendency.

An extension of this principle is the principle of "stepwise motion" used in harmonic progression. Smalley creates a stepwise motion of sustained partials when some dominant partials are changed to values that are close in frequency in subsequent sound events.

If we think of the relationship between speech and song, and extend that idea to ordinary, everyday environmental sounds, then we could say that Smalley is giving voice to ordinary sounds by bringing out the voiced components in those sounds. He achieves this by either reinforcing certain frequencies, transposing the whole sound to a particular pitch, or super-imposing selected sustained frequencies to fuse with those components present in the recorded sound.

**Organisation in time**

The attack-continuant paradigm is a strong structuring force throughout *Wind Chimes*. Pseudo-rhythmic figures are used within the broad attack-continuant framework. Sustained partials are overlapped with the next attack-continuant expansion, providing a "proximity in time" Gestalt glue throughout.
As noted above, simultaneous attacks create "compound sounds" by virtue of their "vertical integration". Nearly-simultaneous attacks create "compound gestures" that are so unique they generate interest and variety.

Repetition is also used in a number of ways. Repeating a sound at short, and ever-decreasing, time intervals produces a feeling of acceleration, propelling the listener towards the goal event. A more subtle technique, used over a wider time-frame, is the creation of a "pre-echo" or the faint anticipation of a sound. Repetition of sound in a new context is also extensively used.

Distinctive sounds make judicious return appearances throughout the piece. To describe the strategic use of these distinctive sounds, we have employed the term “scaffolding”. The skeletal framework is fleshed out through processing and progression. Loud percussive sounds like foot-stomps, struck piano frames, bass drums and so on are used, like punctuation, to mark specific sections.

**Semantic and syntactic progression**

The form of the piece is that it begins with concrete elements which are readily discernible, and progressively moves to an abstract musical space where the spectral and temporal relationships dominate. This is achieved through an increasing use of signal processing, transposition, and an increase in density of sounds as the work progresses. There is a climax in the work and then it relaxes to a point where individual concrete sounds become recognisable once again, right near the end.

When the distinctive concrete sounds, used earlier in the work, reappear in an untreated form at periodic junctures throughout the work, they are generally used as structural signposts. We have used the term “scaffolding” to describe this phenomenon.

Smalley uses a classical Western musical narrative approach with periods of tension and release. On the micro scale, the tension-release pattern is mirrored in the predominant use of the attack-continuant model. On the macro time scale, the compression-relaxation accumulation of percussive short sounds followed by layers of long sustained sounds represents another variation of the tension-release pattern.
Forward motion is provided by connection between successive sounds. Proximity in the frequency domain is coupled with an increasing density of sound event activity - building excitement. These propulsive tendencies are reinforced through the increasing use of signal processing. We can see that schema employed at the micro level reinforce the macro level formal organisation.

7.2.4 Effectiveness of the Framework

Is the framework an effective approach to the analysis of acousmatic music?

The framework developed for the analysis of acousmatic music has been effective in that:

- It provides a systematic reference point for a thorough examination of acousmatic musical works by addressing the detail through the adoption of the SIAM procedure, and by addressing the “big picture” through the use of the listener interpretation-attention notion, which allows the analyst to discuss the work from the point of view of a number of different interpretations (psychological, semiotic, compositional).

- The SIAM framework has resulted in a number of new insights into the particular work Wind Chimes by Denis Smalley. Hopefully this has proven to be valuable information in its own right.

Of course, to be fully tested and refined into a robust methodology, the framework needs to be applied to other acousmatic works, and this would also enable comparative studies to be conducted. Some works may require changes and refinements to the framework that could accommodate more rhythmic pieces, for example. The “listener interpretation” scheme may need to be expanded to provide for phenomenological or philosophical approaches too.

These are some of the issues taken up in the final section on the implications for future research.
7.2.5 Implications for Future Research – or “Unanswered Questions”

Early on in the study it was discovered that there is a paucity of research on the perception and cognition of acousmatic music. In contrast, there has been a lot of research on the perception of Western tonal music and some interesting, although perhaps less voluminous, work on the perception of everyday environmental sounds. The current study has attempted to adapt these two bodies of work to create the SIAM framework, but there is a lot of scope for more experimentation on aspects of the perception of acousmatic music. One critical area for investigation is the role of grouping and hierarchy formation in acousmatic music.

Hierarchy in Acousmatic Music

Many writers have tried to come to terms with hierarchy in music, but it is an especially difficult topic in acousmatic music. James Tenney, while not specifically writing about acousmatic music, attempted to combine Gestalt principles with the articulation of a hierarchical classification system, up to the level of overall form (Tenney, 1988). Tenney’s *Meta+Hodos A Phenomenology of 20th-Century Musical Materials and an Approach to the Study of Form* was first written in 1961, and his *META Meta+Hodos* was written in 1975. They were later combined, revised, and published in a second edition in 1988.

With respect to perceptual organisation, Tenney defined the following terms:

- **Element**: is a *temporal Gestalt unit* (TG) at the lowest hierarchical level. An element is perceived as being singular and not divisible into lower level TGs.

- **Clang**: is a TG at the next higher hierarchical level (*2nd* level).

- **Sequence**: is a TG at the next higher hierarchical level (*3rd* level).

Tenney describes their relationship in the following way:

A clang thus consists of a temporal succession of two-or-more elements; a sequence consists of a temporal succession of two-or-more clangs. Note that a combination of two-or-more elements occurring simultaneously does not necessarily constitute a clang. (Tenney, 1988:101)
A complication immediately arises in the sense that it is possible to perceive simultaneous “elements”, so Tenney identifies four different types of vertical texture: simple-monophonic; simple-polyphonic; compound-monophonic; compound-polyphonic.

In Tenney’s view, hierarchy is determined by cohesion and segregation, which are influenced by proximity and similarity. Other secondary factors he describes as: “accent”, “repetition”, “objective set”, and “subjective set”. (Tenney, 1988:103)

Tenney discusses musical parameters and then moves on to formal perception and description. On the perception of form, Tenney writes:

**PROPOSITION IV:** The perception of form at any hierarchical level involves the apprehension of three distinct aspects of form, at that and all lower levels. These three aspects of form will be called *state*, *shape* and *structure*. (Tenney, 1988:107)

This is not the place to elaborate on state, shape and structure, but Tenney does provide some clues as to how state, shape and structure may form the basis of encoding at different hierarchical levels. Perhaps some experiments on the perception of acousmatic music can be devised using Tenney’s notions as a basis.

A different approach to hierarchy is taken by Van Leeuwen (1999). He writes on perspective, immersion, and social distance. With reference to a number of different types of music, including “soundscapes”, Van Leeuwen makes the following points:

1. The semiotic system of aural perspective divides simultaneous sounds into groups, and places these groups at different distances from the listener, so as to make the listener relate to them in different ways.

2. The sound may either be divided into three groups (positioned as Figure, Ground, and Field) or two groups (positioned as Figure and Ground or as Figure and Field). When there is no perspective, there is only Figure. (Van Leeuwen, 1999:22-23)

Immersion is the opposite to perspective, and Van Leeuwen refers to it as “wrap-around” sound (Van Leeuwen, 1999:28). Sounds seem to come from everywhere at once and he cites examples such as evergreen forests and large churches.
In a real world sound environment, the “social distance” associated with sounds is closely coupled with the above notion of “perspective”. But according to Van Leeuwen:

... the technology of amplification and recording has uncoupled the two, and allowed them to become independent semiotic variables. As a result a soft breathy whisper can now stand out clearly against loud drums or brass sections... (Van Leeuwen, 1999:24-25)

Van Leeuwen then goes on to list five forms of social distance: Intimate; Personal; Informal; Formal; Public. He then provides an illustration of a system network of aural perspective and social distance, and demonstrates how a work, for example a radio play, can be analysed according to such a network.

So, for the purposes of future investigation, does Van Leeuwen’s work on perspective and social distance provide some basis for the design of perception experiments relevant to hierarchy and acousmatic music?

Lerdahl (1987) has also provided some interesting quotes that could form the basis of research questions to be addressed through perception studies of acousmatic music:

If the elements of the novel organisation are arranged according to some foreign principle – say, serial (permutational) operations – is the resulting sequence as learnable as the previous one? (Lerdahl, 1987:156-157)

And

For trees [elaborative hierarchies] of any interest to arise, the continua must be capable of arrangement in arrays of at least two dimensions, thereby creating a “cognitive space”. Within this space, certain moves become possible... (Lerdahl, 1987:157)

What are such “cognitive spaces” for acousmatic music?

Here, further questions can be posed that may inform experimental design of acousmatic music perception studies related to the topic of hierarchies:

- What processes might activate a timbre hierarchy? How can information gained from these processes be combined with other sound parameters to define an event hierarchy?
How do timbral hierarchies combine with pitch and rhythm hierarchies in acousmatic music?

Does hierarchy resolve the conflict between fluidity required for sonority and fixity needed for syntax?

Further Experiments

The following questions may inform experimental design of more general perception studies of relevance to acousmatic music:

- Related to some previous experiments in musical grouping, the segmentation of an acousmatic work could be tested for musicians and non-musicians along the lines of the experiment performed by Deliège and El Ahmadi (1990).

- Is the transformation of timbre like the gap-fill phenomenon found with pitch perception?

- Similarity is one of the prime criteria for categorisation (Barsalou, 1992). Does surrogacy arise out of the natural propensity for categorisation?

- Is there a notion such as timbral dissonance for natural sounds and for abstract sounds? Are there timbral archetypes or prototypes?

- What types of mental models are there for acousmatic music? Experiments could be devised that explore episodic memory, for example, that relate to auditory “scene” representation.

- What non-verbal events can participants remember from a long audio “scene”? The memorisation of a series of random sound events could be compared with a series of semantically-related sound events, for example.

- Is short-term sensing identifying sound sources and attributes, while tracking over longer periods is associated with extracting some “message” from the audio stream?
What are the inter-modal relationships between audition, language and vision? For example, an experiment could be devised where the participants verbalise their self-talk while listening to acousmatic music over headphones. For more on “musical imagery” see Godøy and Jørgensen (2001).

**Patterning, Progression and Prolongation**

Patterning in acousmatic music is associated with sequences of sounds. Progression and prolongation are also tied up with patterning. In order to devise experiments that may cover this area, we need to discuss these concepts a little further.

Sequences can be characterised on the semantic-syntactic continuum. Semantic sequences are of two main types:

1. *Environmentally consistent sequence:* This is a logical connection (or ordering) of the events that is consistent with the sequence of sounds as they occur in the environment. For example, a car door opens; then closes; we hear the sound of the starter motor; the engine idles; and the motor revs higher as the car moves off. We could describe this sequence as perhaps environmentally consistent with the “car sounds” schema.

2. *Contradictory sequence:* The ordering of sound events is not logical when compared with the environment. This may be represented along with two main schemas:

   a. Homogeneous Schema: For example we may employ the “car sounds” schema, but the sequence may be illogical: repeated closing of a car door is followed by the motor revving higher and then the starter motor sounds.

   b. Heterogeneous Schema: We hear a sequence of sounds: a car, a baby, a crowd scene, and there is no logic to the sequence.

A syntactic sequence must generally display some form of sequential integration. As we have seen, such integration is achieved through similarity and/or proximity, either temporally, spectrally or both. Transformation is an important operation as
incremental changes to either the temporal profile or spectral profile will maintain links in the syntactic chain.

A distinction should be made between progression and prolongation. Progression is goal-directed. In acousmatic music it is transformation that assists the feeling of progression. We move from source to target sound via small transformations and there is a feeling of movement away from the source to the destination.

Progression can also be achieved via connection in the frequency domain (spectral profile or spectral register) or the time domain (amplitude profile or temporal proximity). Connection between sequential events via frequency is like the principle of the use of common tones in functional harmony.

In summary it can be noted that syntax within acousmatic music will often be a patterned sequence of events or a related combination of events. A sequence may involve repetition, prolongation or progression. Functional harmonic progression in tonal music can be replaced by functional timbral progression in acousmatic music, through the application of the principle of common frequencies or the principle of transformation.

How these notions might be tested in perception experiments remains fodder for future work.

Implications of Future Research Resulting from the Current Study

The application of the SIAM framework remains to be tested on other acousmatic works. How robust is the framework on other Smalley works, and the whole of the acoumatic music repertoire? Must it be adapted or augmented for use with other works? Will it provide the desired common framework so that comparative studies can be carried out between different works?

Reduction of labour time might be achieved through some data reduction and/or automation techniques. By using models of the human auditory system, the number of frequency bands could be reduced and the estimation of salient pitches could be facilitated through the use of auto-correlation methods. There has been at least one
published attempt at an automated auditory scene analysis method, but the difficulties are enormous (Ellis, 1998).

Modifications to the interactive study score may entail annotation directly on an enlarged sonogram rather than using a separate sound event panel. There may also be some 3-D representations that could be developed in connection with the auditory modelling techniques, mentioned in the previous paragraph, that could prove useful.

What has emerged in this study is a picture of composers like Denis Smalley whose technique is to split open the spectrum of a sound in order to fuse its fragments to other split spectra in a complex web of fusion and fission. The entwined contrapuntal textures that acousmatic music explores are not only tightly bonded in frequency and time, but they show a glimpse of reality and semantic meaning only to whisk it away in an instant into a fanciful world of pure sound as enveloping as the sounds that a developing child may experience within its mother’s womb.

In drawing this present study to a close, I am moved to quote the final stanza from an article by the eminent Leonard B. Meyer:

I am, appropriately, uncertain about how to end this essay – and that makes me somewhat sad. Yet, at the same time and for the same reason, I feel glad – glad because connecting puzzled uncertainty with emotion seems to confirm my basic argument. (Meyer, 2001:359)
Bibliography


HEVNER, K. (1936) Experimental studies of the elements of expression in music. 


Appendix 1 – Composition Folio

The composition “folio” consists of a set of electroacoustic music compositions on the accompanying audio CD. The five compositions make up a suite called C’est la vie which explores various aspects of life. The five pieces that make up C’est la vie are:

1. *La vie naturelle* - Natural Life
2. *Maison* - Home life
3. *Artificier artificiel* - Artificial Life
4. *Travail* - Working Life
5. *Mon Dieu* - Religious Life

The following text provides some explanatory notes on each piece.

**La vie naturelle**

Duration: 10’ 10”

In *La vie naturelle*, I have tried to create a hyper-reality where the sounds of nature are intensified into their own sound world. A world where the sounds take on a life of their own, detached from our day-to-day experience. We are led slowly into that world, we briefly witness the drama there, and then we are eased back to reality.

The source sounds were recorded in Taylor Park, Torquay, Victoria. The principle techniques used in this work are extreme stretching in the time domain (independent of frequency) and transposition of pitch (independent of time). The main software tools used were *Cecilia* (a graphic front end to the *Csound* software synthesis language) and *Csound* itself.

**Maison**

Duration: 14’ 28”

*Maison* is the second movement of the C’est la vie suite. *Maison* takes home life, or family life, as its primary source material. Toys “taking over the household” is an age
old theme that story-tellers have used for centuries. What happens when ordinary household sounds take on a life of their own?

*Maison* makes use of many contraptions created in Ross Bencina's *Audiomulch* program (see www.audiomulch.com ). Original source sounds include kitchen sounds, hallway sounds, and bathroom sounds (created by Caitlin and James Hirst).

**Artificier artificiel**

Duration: 4’ 20”

Literally the artificial pyrotechnist, *Artificier artificiel* explores artificial life. Although the sounds used are recordings of ordinary, everyday sounds, the work draws upon the flight patterns of a flock of artificial birds. The sound component derives from a 2-dimensional landscape of sounds laid out in a ‘semantic grid’. As the birds fly over this landscape, a single bird may ‘see’ a sound-making object – a cow or a car for example. If it does, it will ‘hear’ a cow sound or a car sound. There are 10 birds, but 5 of them are deaf, so only up to 5 sounds at any one time can be heard by the listener. There are 50 sound ‘objects’ spread over the landscape area.

*Artificier artificiel* is a laptop performance work, with sounds and images created in real-time using the programming language *PD* with the *GEM OpenGL* graphic library. The on-screen image displays spherical shapes representing the flight path of each bird in the flock. Using this visual cue, the performer can manipulate various parameters to vary the paths of the birds, so they flock together or they avoid each other and go off into different directions, or there is a combination of behaviours – all under the performer’s control. The speed that the birds travel across the aural landscape can also be varied.

This version is a fixed concert version of the piece, and was recorded in real-time in a single pass. The result is a collage of semantic associations, with rhythmic juxtapositions. The semantic grid was drawn up according to the MDS results presented by Gygi (2001) – see Figure 4.2.
**Travail**

Duration: 14’ 35”

Pronunciation: tr&-'vA(&)l, 'tra-"vAl

Function: noun

Etymology: Middle English, from Old French, from travailler to torture, labor, from (assumed) Vulgar Latin trepaliare to torture, from Late Latin trepalium instrument of torture, from Latin tripalis having three stakes, from tri- + palus stake

*Travail* explores all of these definitions through sound, as an expression of the trials and tribulations of working life. *Travail* also makes use of contraptions created in Ross Bencina's *Audiomulch* program to process the source sounds - which originated from recordings of an espresso coffee making machine in action (coffee being a good antidote to work).

**Mon Dieu**

Duration: 8’ 44”

*Mon Dieu* is the final movement from *C'est la vie*. The main elements of the work were developed and organised using some of Carl Jung's ideas, e.g. Progression and Regression - Energy flows, the psyche is never in equilibrium long. Something comes along to disturb it and thus progression and regression are required. This is a little like tension and release in music. Stasis is disturbed by some new element, which is then assimilated into the work.

The work makes extensive use of synthesis of the singing voice. It uses the FOF vocal formant module by Jean Piché, which is distributed in his *Cecilia* front end for the *Csound* software synthesis and signal processing language. Fragments of Josquin's *Missa Pange Lingua* are used to provide references to Western religion in the first part of the work, while Eastern modes are used in the latter part of the work. Other psychoanalytic archetypes make appearances, such as the sounds of a tennis match, rowing, and a cricket match.
These Jungian psychoanalytic associations of sport are mixed with religious psychoanalytic associations from the composer's youth, and progressing through changes in life. The image of strolling by the river, watching rowers, hearing a tennis match next to the church while choral rehearsals are under way also has some connection to the theme of the work. In the middle, the work fractures as influences from Asian philosophy then take hold and it concludes with the expression of an Eastern melodic line as the meditation comes to a point of closure.
Appendix 2 – Supplementary CD-ROM

Interactive Study Score

The second CD in this submission is a data CD containing the interactive study score of Denis Smalley’s *Wind Chimes*. It is designed to run as a Flash-based interactive, within an internet browser program such as Internet Explorer or Mozilla Firefox.

Instructions for Use

- Insert the CD into your computer.

- Open the folder called “wind_chimes_study_score”.

- Double click on the file called “ReadMe.html”, or open it from within your internet browser program. Click on the image shown on the web page, and the interactive study score should open in a new window. If you have any blocking protection in your browser program, you should allow this interactive content.

- Turn up the audio volume on your computer and then press the “play” button. You will hear the audio, and a vertical red line will indicate the current playback position within the study score.

The player has the following controls:

- Play/Pause button.

- Return to start button.

- Progress bar with a button that you can use to move forward & backward within the piece.

You will also need to have the latest Flash plugin (Flash 6.0 or later) installed in your browser – available from http://www.macromedia.com
Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
HIRST, DAVID

Title:
The development of a cognitive framework for the analysis of acousmatic music

Date:
2006-05

Citation:

Publication Status:
Unpublished

Persistent Link:
http://hdl.handle.net/11343/39067

File Description:
The development of a cognitive framework for the analysis of acousmatic music

Terms and Conditions:
Terms and Conditions: Copyright in works deposited in Minerva Access is retained by the copyright owner. The work may not be altered without permission from the copyright owner. Readers may only download, print and save electronic copies of whole works for their own personal non-commercial use. Any use that exceeds these limits requires permission from the copyright owner. Attribution is essential when quoting or paraphrasing from these works.