DYNAMIC RERESENTATION

OF

MUSICAL STRUCTURE

By

Edward Wilson Large, Ph.D.

The Ohio State Universtiy

1994

Copyright by

Edward W. Large

1994

LIST OF TABLES

TABLE . . . . . PAGE

1. Squared correlation coefficients for theoretical predictions and improvisation-based measures. 25
2. Squared correlation coefficients for network reconstructions. 48

LIST OF FIGURES

FIGURE . . . . . PAGE

1. Representational formalisms for music perception and cognition. 3
2. A rhythmic grouping for *Hush little baby*.. 10
3. Melodies and variations. *Hush little baby* (top), and *Mary had a little lamb* (bottom), showing (A) subject melodies, and (B) improvised variations on the subject melodies. 12
4. Analysis of *Hush little baby* following Lerdahl & Jackendoff (Lerdahl, & Jackendoff, 1983). The original melody is shown in musical notation and the brackets below mark the time-span segmentation. Solid brackets describe the contribution of grouping structure (Section 2.1); dotted brackets describe the contribution of metrical structure (Section 2.3). The tree above the notated melody is a time-span reduction. The lower staves show the dominant events for each level of the time-span segmentation.. 13
5. The first three levels of reduction for *Hush little baby* and its variation (from Figure 3). The reductions are identical at the third level. 14
6. A metrical structure for *Hush little baby*.. 17
7. Metric strata. Simple ratios imply grouping. Polyrhythmic ratios do not imply grouping. 18
8. Analysis of *Hush little baby* showing metrical structure, time-span segmentation, and time-span reduction. The quantifications of relative importance for each event are shown below the segmentation. 25
9. *Mary had a little lamb*: (A) musical notation, (B) onsets times prescribed by score, (C) discrete Fourier transform of B, (D) auto-correlation function of B, (E) discrete Fourier transform of D. 28
10. Transcription of an improvisation on *Hush little baby*: (A) musical notation, (B) onsets time prescribed by score, (C) discrete Fourier transform of B, (D) auto-correlation function of B, (E) discrete Fourier transform of D.. 29
11. Performance of *Mary had a little lamb*: (A) musical notation, (B) onset times recorded in the performance, (C) discrete Fourier transform of B, (D) auto-correlation function of B, (E) discrete Fourier transform of D. 30

12. Improvisation on *Hush little baby*: (A) musical notation, (B) onset times recorded in the performance, (C) discrete Fourier transform of B, (D) auto-correlation function of B, (E) discrete Fourier transform of D.. 31
13. Bar plot showing a significant two-way interaction of subject and melody in deviation from average tempo. 32
14. Encoding and decoding of a musical sequence by a RAAM network. (A) Based on a vector representation for each event and a constituent structure analysis, the compressor combines the group (a b) into a single vector, R1, (c null) into the vector R2,and then combines (R1 R2) into the vector R3. (B) The reconstructor decodes the vector R3 to produce (R1' R2'). It then decodes R1' to produce the facsimile (a' b') and R2' into (c' null').. 35
15. Buffering scheme for encoding either duple and triple grouping structures. (A) Three buffers cannot discriminate between a group of two events and a group of three events in which the middle event is a rest. (B) Four input buffers can make the discrimination.. 39
16. Network reconstructions of *Mary had a little lamb* in Experiment 1.. 41
17. Network reconstructions of *Baa baa black sheep* in Experiment 1. 42
18. Network reconstructions of *Hush little baby* in Experiment 1. 42
19. Simplified schematic of modular RAAM encoding used in experiment 2. (A) In the encoding diagram, time flows from left to right and bottom to top. (B) In the decoding diagram, time flows from top to bottom and left to right. The network determines when to stop decoding automatically. 44
20. Original melodies and network reconstructions: (A) *Mary had a little lamb* (Known), (B) *Baa baa black sheep* (Variant), (C) *Hush little baby* (Novel). Each melody was reconstructed from several codes (the half-note level RAAM), and from a single code (the whole-tune level RAAM). X's denotes failures in network reconstructions. 46
21. A periodic signal and the response of an integrate-and-fire oscillator. 64
22. An input signal to the oscillator model 68
23. Output pulses (temporal receptive fields) for two different values of τ. (A) $\tau = 0.10$, (B) $\tau = 0.05$. τ measures the width of the temporal receptive field. 69
24. The effect delta rules for phase and period given in Equations (3) and (4) for two different values of τ, (A) $\tau = 0.10$, (B) $\tau = 0.05$. This figure illustrates how τ gives the amount of variability that the unit will tolerate input signal. 71
25. The relationship between Ω and τ, according to Equation 6.. 72

26. The effect of the delta rule for variability (τ) given in Equation 7 for two different values of τ. (A) $\tau = 0.10$. (B) $\tau = 0.05$. The y-axis gives $\Delta\Omega$ values, and τ is calculated from Ω according to Equation 6. 73
27. A possible relationship between Ω and c, according to Equation 8. This provides a measure of performance. 74
28. Single unit tracking a periodic signal: (A) input signal, (B) oscillator output pulses, (C) oscillator period, (D) τ. 75
29. Single unit tracking a periodic signal: (A) input signal, (B) oscillator output pulses, (C) oscillator period, (D) τ. 77
30. The state space for a system of two oscillators, a torus. A position on the surface of the torus describes a the combined system as a pair of phases. 79
31. A graph of the finite difference equation given by Equation 9 for $p/q = 2/3$, and $b = 0$. 80
32. An Arnol'd tongues diagram (A) and the Farey tree (B). 81
33. Stimulus impulse times, oscillator expected onset times and phase tracking delta rule. Solid lines show the situation up to time *Tj*, dotted lines show the situation after time T_j: the impulse causes a change in driven oscillator cycle length. 83
34. An empirical regime diagram for the phase-coupled model with $\gamma = 0$. 85
35. An empirical regime diagram for the phase-coupled model with $\tau = 0.10$. 86
36. An empirical regime diagram for the phase-coupled model with $\tau = 0.05$. 87
37. An empirical regime diagram for the period-coupled model with $\gamma = 0$. 89
38. An empirical regime diagram for the period-coupled model, with $\tau = 0.10$. 90
39. An empirical regime diagram for the period-coupled model with $\tau = 0.05$. 91
40. An oscillator tracking the rhythm of *Baa baa black sheep* (rubato = 0.05, $\overline{|\phi'|} = 0.08$; $R^2 = 0.34$, $p < 0.05$). 96
41. An oscillator tracking the rhythm of *Hush little baby* (rubato = 0.07, $\overline{|\phi'|} = 0.04$; $R^2 = 0.32$, $p = 0.29$). 99
42. An oscillator tracking the rhythm of *Hush little baby* (rubato = 0.07, $\overline{|\phi'|} = 0.20$; $R^2 = 0.39$; $p = 0.18$). 101
43. Oscillator tracking an improvisation on *Mary had a little lamb* (grace notes are not transcribed); rubato = 0.25, $\overline{|\phi'|} = 0.17$, R^2=0.04, p = 0.87). 103

44. Oscillator tracking an improvisation on *Hush little baby* (rubato = 0.25, $\overline{|\phi'|} = 0.17$; R^2=0.04, p = 0.87). 104
45. Oscillator tracking an improvisation on *Baa baa black sheep* (not transcribed; rubato = 0.39, $\overline{|\phi'|} = 0.20$; R^2=0.20, p = 0.09). 106
46. Four test case for Experiment 1: two simple ratios, and two polyrhythmic ratios. 107
47. Response of Oscillators 1 and 3 to a 2:1 rhythm. 108
48. Response of Oscillators 1 and 4 to a 3:1 rhythm. 108
49. Response of Oscillators 1 and 2 to a 3:2 polyrhythm. 109
50. Response of Oscillators 1 and 2 to a 4:3 polyrhythm. 109
51. Confidence of each oscillator and rhythm. Each curve is marked with the period of the corresponding oscillator. Those that do not acquire stable mode locks are indicated with '?'. 110
52. The response of two oscillators to rhythmic input from a digital sample of read speech (from Cooper & Meyer, 1960). The digital sample was filtered using an edge-detection algorithm, and the results of the edge detection algorithm were used as input. The input impulses (dotted lines) and output beats (solid lines) are shown for Oscillators 1 and 2 in panels A and C, respectively. The periods of Oscillators 1 and 2 are shown in panels B and D, respectively.. 117

TABLE OF CONTENTS

LIST OF TABLES ii

LIST OF FIGURES iii

CHAPTER PAGE

I. MENTAL REPRESENTATIONS FOR MUSIC 1

1.1 Introduction. 1
1.2 Method and Research Plan 3
1.2.1 Computing Mental Representations for Music 3
1.2.2 Empirical Measures of Model Performance 4
1.3 General Relevance 5
1.3.1 Connectionist Temporal Sequence Processing 5
1.3.2 Speech 6
1.3.3 Motor Coordination 7
1.4 Outline of the Thesis. 7

II. SEQUENCE STRUCTURE AND TEMPORAL STRUCTURE IN MUSIC. 9

2.1 Rhythmic Grouping 9
2.2 Patterns and Reductions 10
2.3 Dynamic Attending, Temporal Expectancy, and the Entrainment Hypothesis 14

III. A STUDY OF MUSIC PERFORMANCE AND IMPROVISATION 19

3.1 Structural Relationships Among Sequence Events. 19
3.2 Temporal Structure in Music Performance and Improvisation. 21
3.2.1 Method. 22
3.2.1.1 Subjects. 22
3.2.1.2 Materials 22
3.2.1.3 Apparatus 22
3.2.1.4 Procedure 23
3.3 Analysis #1: Mental Representation of Melodies. 23
3.3.1 Coding Improvisations. 23

3.3.2 Comparison with Theoretical Predictions 24
3.4 Discussion 26
3.5 Analysis #2: Performance and Improvisation Timing 26
3.5.1 Timing Analyses 26
3.5.2 Rubato 32
3.6 Discussion 33

IV. COMPUTING REDUCED MEMORY REPRESENTATIONS 34

4.1 Connectionism and Reductionist Music Theory 34
4.2 A RAAM Architecture for Music 37
4.3 Experiment 1: Balanced Tree Structures 39
4.3.1 Training 39
4.3.2 Testing 40
4.4 Experiment 2: Unbalanced Tree Structures 43
4.4.1 Training 43
4.4.2 Testing 43
4.4.2.1 Tests of Well-Formedness 45
4.4.2.2 Tests of Representational Structure 47
4.5 Discussion 48

V. SEQUENCE PROCESSING AND TEMPORAL PROCESSING 51

5.1 Temporal Sequence Processing 51
5.1.1 Sequence Processing 52
5.1.2 Temporal Processing 56
5.1.2.1 Temporal Sequence Processing in Relative-Time 56
5.1.2.2 Temporal Sequence Processing in Real-Time 58
5.1.2.3 From Real-Time to Relative Time 60
5.2 The Computation of Temporal Structure 61
5.2.1 Quantization and Time-Warping 61
5.2.2 Structural Analysis of Rhythmic Signals 62
5.2.3 Dynamic Processing of Input Rhythms 62

VI. SYNCHRONIZATION TO COMPLEX SIGNALS 66

6.1 Definitions 67
6.1.1 Synchronization and Entrainment 67
6.2 The Oscillator Model 68

6.2.1 Output Events 68
6.2.2 Phase-tracking and Period-tracking 70
6.2.3 Tracking Variability 72
6.2.3.1 Confidence 73
6.3 Oscillator Behavior 74
6.3.1 Compound Units 76
6.4 Discussion 76

VII. MODELING BEAT PERCEPTION AS A DYNAMICAL SYSTEM 78

7.1 The Sine Circle Map 78
7.2 Adapting the Circle Map 82
7.2.1 Phase-Coupling 82
7.2.2 Period-Coupling 88
7.3 An Efficient Algorithm 92
7.4 Discussion 93

VIII. SOME EXPERIMENTS WITH THE OSCILLATOR MODEL 94

8.1 Performed Musical Rhythms 94
8.1.1 Performance of Notated Melodies 98
8.1.1.1 Case 1 98
8.1.1.2 Case 2 100
8.1.2 Performance of Improvised Variations 100
8.1.2.1 Case 3 102
8.1.2.2 Case 4 102
8.1.2.3 Case 5 105
8.2 Discussion 105
8.3 The Perception of Metric Relationships 107
8.4 Discussion 111

IX. IMPLICATIONS: MUSIC COGNITION AND BEYOND 113

9.1 The Basic Findings 113
9.1.1 Computing Structural Descriptions for Musical Sequences . . 113
9.1.2 Dynamic Representation of Temporal Structure 114
9.1.3 Sequence Structure and Temporal Structure 114
9.2 General Significance 115
9.2.1 Connectionist Temporal Sequence Processing 115

9.2.1.1 Oscillation and Synchronization in Dynamic Feature Binding Networks 115
9.2.2 Speech 116
9.2.3 Motor Coordination 118
9.3 Future Work 118
9.4 Closing Thoughts 120

LIST OF REFERENCES 121

CHAPTER I

MENTAL REPRESENTATIONS FOR MUSIC

1.1 Introduction

The problem of how the human brain perceives and represents complex, temporally structured sequences of events is central to cognitive science. The basic questions of temporal sequence processing recur throughout the study of human activity, in domains as diverse as language, vision, and motor coordination. Understanding how temporal sequences may be coded as patterns of activation in neural networks is a central issue. This dissertation addresses two questions that are important in understanding the representation of structured sequences. The first regards the *acquisition and representation of structural relationships among events*. Sequence structure is important in representing sequences with long distance temporal dependencies, and in learning structured systems of communication. The second regards the *representation of temporal relationships among events*. Temporal structure is important in recognizing and representing sequences independent of presentation rate, while retaining sensitivity to relative timing relationships. These two issues are intimately related, and this dissertation addresses the nature of this relationship.

Music provides a domain that is in many ways ideal for the studying perception and representation of complex temporal sequences. Music is a highly structured form of communication requiring knowledge that is shared among composers, performers, and listeners. Unlike other forms of communication, music is an activity in which communication appears to take place without explicit referential semantic content. Thus music provides a domain in which to study issues of temporal sequence processing without positing a great deal of extra-musical knowledge. At the same time music provides a rich source of data, generated by a natural human activity, in which complex sequential and temporal relationships abound.

The issue of structural relationships among sequence elements figures prominently in many accounts of mental representation for musical structure. For example, in a particular musical context, certain events may be perceived as relatively important, while others are perceived as mere elaborations of more important events (Lerdahl, & Jackendoff, 1983; Schenker, 1979). The notion of relative importance in music is roughly comparable to the notion of prosodic structure in language. Knowledge of sequence structure figures prominently in accounts of how people learn musical style systems, and how people recognize musical variation. Depending on the musical dimension(s) under consideration the nature of the description varies, but each relies on some abstract system of knowledge representing underlying sequential regularities. Knowledge for

creating structural descriptions may be innate, or it may reflect the statistical regularities of a particular musical culture or style (e.g. Knopoff & Hutchinson, 1978; Palmer & Krumhansl, 1990; Mozer, 1994).

The dynamic, or time-varying aspect of music perception and cognition is also prominent in theoretical accounts of musical structure. Such accounts often emphasize the ability of listeners to anticipate upcoming events from what has gone before. Sequential accounts of musical expectancy focus on *what* events a listener expects to occur, and are often expressed in terms of learned probability relationships that describe the statistical regularities of sequences (e.g. Bharucha & P. Todd, 1991; Meyer, 1956; Mozer, 1994). Temporal accounts of expectancy focus explicitly on the question of *when* listeners expect events to occur (e.g. Jones, 1981b; Large & Kolen, in press). This distinction will prove useful in understanding the temporal sequence processing properties of artificial neural networks. Sequence processing systems must deal with the serial ordering of future events using knowledge of sequential structure. Temporal processing systems must cope with the question of when events are likely to occur by exploiting knowledge of temporal structure.

One way to think about the problems of representing structure in music cognition is according to the types of representational formalisms employed by theorists working on different problems. Figure 1 shows one way to do this. At the bottom of Figure 1 is a representation of an acoustic flow as a digital signal, a time-series of amplitudes describing the musical surface. This formalism is useful for studying the perception of pitch, loudness, and timbre. It is also useful for studying how listeners identify more-or-less discrete musical events such as notes and chords within in the continuous acoustic flow. The next representation abstracts away from the musical surface to describe music as a time-series of discrete events with properties such as pitch, amplitude, timbre, and location in time. This formalism, sometimes called piano-roll notation, is useful for studying the perception of tonal and temporal relationships. The next level abstracts away from the musical surface even further, illustrating ways of describing rhythmic relationships in a musical signal. One is a representation in terms of groups, or chunks, providing a segmentation of the stimulus into nested constituent structures. The second representation describes the onset of events with respect to a metrical grid, making explicit relative time relationships among events. Representation in terms of grouping and meter is necessary to create a familiar type of musical representation, the musical score. Representations of temporal properties are useful in understanding the acquisition and representation of abstract knowledge of musical style. Finally, the tree atop the musical score illustrates representation in terms of structural relationships among events. Structural relationships may be expressed in terms of grammars, schemata, or prototypes, for example. Such representations are useful for characterizing musical style systems, and for understanding creative activities such as musical composition, improvisation, and performance.

Sometimes illustrations such as this are supposed to provide a flow chart of information-processing stages in music perception, performance, and cognition. Read from bottom to top, for example, Figure 1 could be interpreted as a stage model of music perception; read from top to bottom, it could be interpreted as a stage model of music performance. At this point in the understanding of the perception and representation of musical sequences, it is perhaps wise to

avoid such an interpretation. However, studies in music cognition can often be described as attempts to understand the transformation of one representation into another.

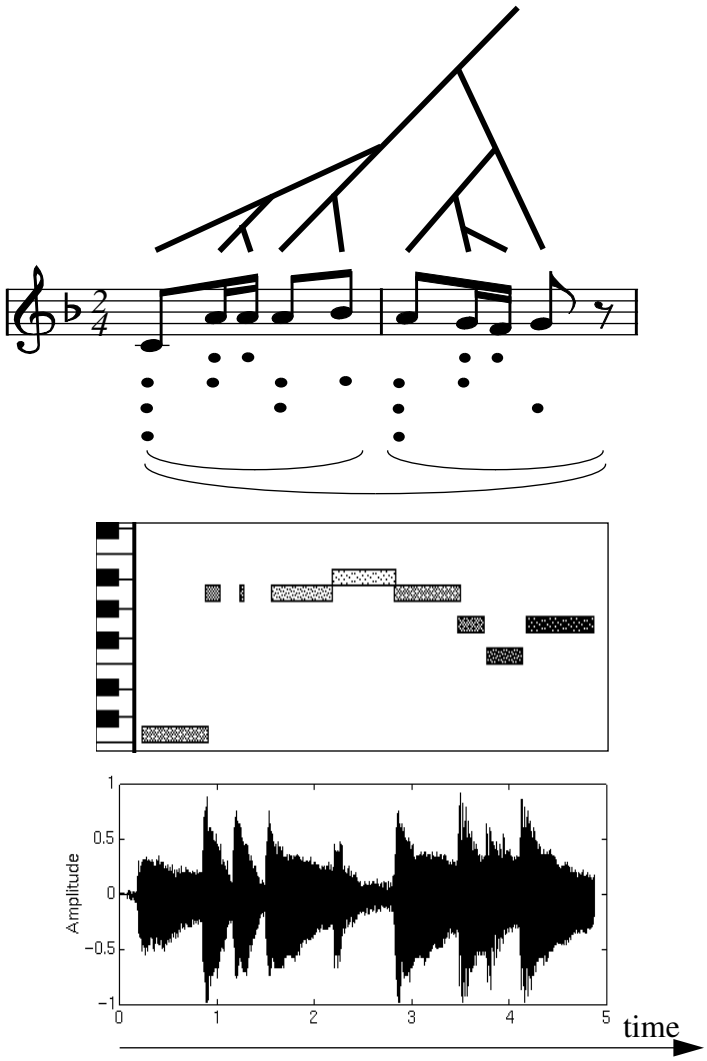


Figure 1: Representational formalisms for music perception and cognition.

1.2 Method and Research Plan

1.2.1 Computing Mental Representations for Music

This dissertation describes two research projects that address separable, but closely related problems. The first study models the acquisition and representation of structural relationships among events in musical sequences, addressing issues of style acquisition and musical variation. With respect to Figure 1, the model takes as input a score-like representation (rhythmic properties are known) and produces a structured description (top of Figure 1). The second study models the perception and representation of temporal relationships among events. With respect to Figure 1, the

model takes as input a piano-roll representation and produces a metrical grid (shown below the score in Figure 1).

The first model addresses the problem of producing structural descriptions for musical sequences. A neural network encodes the rhythmic organization and pitch contents of simple melodies. As the network learns to encode melodies, structurally more important events dominate less important events, as described by reductionist theories of music (Lerdahl, & Jackendoff, 1983; Schenker, 1979). Reductionist theories posit that an experienced listener assigns to a musical sequence a relative importance structure based on previously acquired information: information that is not necessarily present in the individual musical sequence. The network displays a form of learning, providing an example of how listeners may acquire intuitive knowledge through passive exposure to music that allows them to construct reduced memory representations for musical sequences.

The connectionist network successfully captures structural relationships among events by exploiting knowledge about relative timing relationships. This requires the type of information about temporal structure made explicit by the metrical grid of Figure 1. Metrical structure describes an important part of musical phenomenology, the sense of alternating strong and weak beats that accompanies the experience of listening to music (Lerdahl, & Jackendoff, 1983). The basic ability that affords the perception of metrical structure is the ability to perceive the *beat* of a musical sequence.

The second model addresses the perception of temporal structure in musical sequences, specifically the perception of beat and meter. This approach is inspired by dynamic attending theory (Jones, 1976; Jones & Boltz, 1989). Dynamic attending theory describes rhythm perception as a dynamic process in which the temporal organization of rhythm synchronizes, or entrains, a listener's attention. This dissertation describes an entrainment model appropriate for modeling the perception of beat and meter in music. An oscillator tracks the phase and period of periodic components of complex rhythmic patterns, resulting in dynamical system model of beat perception. The self-organizing response of a group of oscillators embodies the perception of metrical structure.

1.2.2 Empirical Measures of Model Performance

Computational modeling of cognitive phenomena can be quite valuable, but only if the behavior of the model can be properly evaluated. Deciding how to judge the behavior of a musical model may the most difficult issue facing computational modeling of musical activity. The models presented in this dissertation will be evaluated with respect to a single data set, collected in an empirical study of music performance (Large, Palmer, & Pollack, 1991; in press). In this study, musicians performed three melodies, and improvised a set of variations on the melodies.

There is an important reason performance data provides a good test for these computational models. Studies of music perception generally require simplifying assumptions. For example, stimuli may be constructed and presented to conform to such strict statistical controls that the

stimuli to which subjects respond is not really music (Butler, 1992). Alternatively, subjects respond to actual segments of music, but only broad responses are measured such as judgements of similarity (Large, 1992; Serafine, Glassman, & Overbeeke, 1989). I do not mean to underestimate the importance of such studies, or suggest that sophisticated perceptual paradigms do not exist in music cognition. However, the computer models presented in this dissertation provide detailed predictions regarding perception. No small set of strictly perceptual studies can provide adequate tests of the models' behavior. The performance data collected, along with simple assumptions about the nature of perceptual and cognitive processes, allowed adequate tests of behavior for these simulations.

To evaluate the model of musical structure, musicians' improvised variations were analyzed to determine aspects of the pianists' mental representations for these three melodies. Improvisations were compared with predictions of structural importance based on reductionist accounts. The evidence from improvisational music performance addresses the validity of reductionist claims and their relationship to the problem of musical variation. It also provides empirical data with which to compare the performance of the connectionist network for producing reduced memory representations.

To evaluate the model of beat perception, both melodic performances and improvisations were analyzed to determine their timing properties, and then used as test cases for the computer model. This analysis differs from previous studies of performance timing in that it does not attempt to determine the significance of deviation from temporal regularity common in musical performance (cf. Palmer, 1988). It assumes instead that the basic job of the beat and meter perception is to track the temporal structure of performances *in spite of* performance timing deviations (cf. Longuet-Higgins & Lee, 1982). The model is evaluated with respect to its performance on this task.

1.3 General Relevance

1.3.1 Connectionist Temporal Sequence Processing

Within the artificial neural network community, a great deal of attention has focused upon questions of temporal sequence processing. One issue researchers have addressed is the ability of short term memory structures to adequately make sequence history available to network processing. This is important in representing sequences in which relationships among events span long temporal intervals and involve high-order statistics (de Vries & Principe, 1992; Jordan, 1986; Wang & Arbib, 1993; for reviews see Mozer, 1993; Wang, in press a). Another issue regards the design of processing strategies and training algorithms that make generalizations about sequence structure that are relevant in particular domains. This is important, for example, in learning the structure of naturally or artificially generated languages (e.g. de Vries & Principe, 1992; Giles, et. al. 1990; Kolen, 1994; Pollack, 1988; Pollack, 1991; Cleeremens, Servan-Schreiber, & McClelland, 1989). These issues mainly involve sequence structure. Time enters the picture, but in a limited way: as a constraint on the maintenance of sequence history in short term memory.

An equally important set of questions regards how systems handle temporal structure. One issue regards the design of systems that are rate-invariant while maintaining sensitivity to relative timing relationships. Systems for processing music and speech, for example, must process sequences independent of absolute presentation rate, yet maintain sensitivity to certain relative time relationships. These issues are related to a problem known as the *quantization*, or *time-warping* problem. The time-warping problem is the problem of deciding what relative-time relationships should be maintained, and what other aspects of timing should be disregarded.

A third set of questions regards how temporal structure and sequence structure should interact in temporal sequence processing. Artificial neural networks, for example, may be designed to exploit temporal structure. The work reported in this dissertation bears directly upon these issues. The neural network model of Chapter IV, learns to represent complex musical sequences. It makes musically and psychologically relevant generalizations about sequence structure. It accomplishes this using a short term memory design that exploits knowledge of temporal structure. The entrainment model of Chapters VI-VIII proposes a way of providing such information about temporal structure, so that any neural network may exploit relative timing information for temporal sequence processing. The relationship between the two models is discussed, and directions for future research are suggested.

1.3.2 Speech

Music perception is interesting, in part because it is related to the perception of speech. For example, the neural network model for learning musical sequence structure is related to theories of prosody in natural language. The network's representations of musical sequences differentially weights musical events as described by reductionist theories of music (Lerdahl, & Jackendoff, 1983; Schenker, 1979). Reductionist theories share a common framework with rule systems for prosodic stress and prominence in speech (Liberman 1975; 1977; Lerdahl, & Jackendoff, 1983; Selkirk, 1978; 1980). Thus, findings of this simulation may be relevant to linguistic models. Design of this network may suggest network designs for learning prominence relationships in spoken language.

There is also a close relationship between theories of metrical structure in music and theories of metrical structure in language (Lerdahl, & Jackendoff, 1983). Lerdahl and Jackendoff, however, note an important difference between the notion of a regularly timed metrical structure in music and the apparent flexibility of timing in natural language. Regular timing (isochronic organization) has not been widely observed in natural language either in the acoustic or in the articulatory domain, yet systematic deviation from isochrony has been shown to reliably communicate linguistic information to listeners. This seemingly paradoxical result has led some researchers to the conclusion that isochrony in language is a perceptual phenomenon (Lehiste, 1977). One possible explanation is that certain units of speech (syllables, stressed syllables, or mora) are approximately regularly timed, but variance in timing constraints introduced by segmental variation (production of different syllables) masks this regularity (e.g. Fowler, 1983; Kelso, et. al., 1985). Results presented in Chapter III show that naturally performed music is not regularly timed either, at least not in any easily measurable way. Nevertheless, the model of beat

perception in music (Chapters VI-VIII) finds temporal regularities, or perceptual isochrony, at multiple time scales. This approach may be applicable to speech timing, and the final chapter presents an example to suggest a way in which the model might be applied to speech.

1.3.3 Motor Coordination

There is a large body of theoretical and empirical work that relates to timing in musical performance. Musical rhythms performed by skilled musicians show deviations from timing regularity (as prescribed by the musical score) that are systematically related to the musical intentions of performers (Sloboda, 1983; Clarke, 1985; Shaffer, Clarke & N. Todd, 1985; N. Todd, 1985; Palmer, 1988; Drake & Palmer, 1993). It is generally assumed that listeners can respond to these perceptual cues and comprehend the intentions of performers, thus deviation form *ideal* timing in musical performance communicates musical information. This suggests a representation of musical timing in two parts, a canonical motor program giving *ideal* durations (as found in a musical score), and a curve that represents deviation from *ideal* timing (e.g. Clarke, 1993). In Chapter V, I define the function of meter perception in complimentary terms – to reverse engineer a motor program that would recreate the rhythm. This approach is consistent with Fowler's (1990) view that the object of auditory perception is the sound-producing source, not the sound itself.

Of broader interest to the motor coordination community is the issue of the representation of motor control structure: hierarchically nested motor programs vs. non-linearly coupled oscillators. The experimental literature on motor coordination reveals many activities, including rhythmic hand movements and cascade juggling, to be consistent with mathematical laws governing coupled oscillations (e.g. Kelso & deGuzman, 1988; Schmidt et. al., 1991; Treffner, & Turvey, 1993; for a review of recent models see Beek, Peper, & van Wieringen, 1992). Shaffer (1981) has proposed that the performance of two-handed polyrhythms in music may be described as the entrainment of clocks. However, this issue remains unresolved. The model of meter perception that I propose is based on the notion of non-linearly coupled oscillations. I will argue that the success of this model in the perceptual domain also provides support for its use in the motor domain.

1.4 Outline of the Thesis

Chapter II provides a more detailed background of music cognition, focussing on two issues that will be of concern in the modeling efforts of subsequent chapters: representation of structural relationships and representation of temporal relationships among events in musical sequences. Theoretical and empirical work pertinent to these topics is reviewed.

Chapter III describes the collection and analysis of a data set that will be central to the material in following chapters. A study of musical performance and improvisation was conducted in which musicians performed simple melodies from musical notation, and then improvised variations on these melodies. The data will be used to evaluate the performance of the models described in the following chapters. First improvisations are collected and used to assess the structure of the performers' mental representations for these melodies. An analysis comparing

improvised variations with the predictions of a reductionist theory (Lerdahl, & Jackendoff, 1983) provides support for the theory, and provides predictions to test the computer simulation described in Chapter IV. Next, both performances and improvisations are analyzed with respect to the timing of the performances. Methods of Fourier analysis and auto-correlation are used to assess timing properties of the performances, and then deviation from "ideal" timing is assessed using a measure of performance rubato. These materials will be used to test a model of beat perception in Chapter VIII.

Chapter IV describes a connectionist model of structural representation for musical melodies. Memory representations for musical sequences are modeled using recursive distributed representations (Pollack, 1988; Pollack, 1990), a connectionist formalism that allows the representation of symbolic data structures as patterns of activation in connectionist networks. A computational experiment is described in which a neural network is trained to produce recursive distributed representations for the three melodies used in the improvisation study (Chapter II). An examination of the reduced descriptions reveals that the representations differentially weight musical events, emphasizing some aspects of the musical content over others. Thus, the network captures a theoretically important type of structural relationship among sequence events. Results are compared with the empirical study to address whether the network's differential weightings agree with the relative importance of events inferred from the improvisational music performances. Some of the material in Chapters Two and Three appear in (Large, Palmer, & Pollack, in press).

Chapter V begins a change in perspective. Results of the previous simulation are assessed, and assumptions about temporal structure are made explicit. The distinction between sequence processing and temporal processing is explored, and temporal sequence processing architectures for music are reviewed with regard to this distinction. The chapter then reviews previous approaches to analyzing temporal structure in music.

Chapter VI proposes an entrainment model of the process of tracking beats in musical sequences that is appropriate for modeling aspects of human rhythm perception. Chapter VII uses the entrainment model to derive a dynamical system model of beat perception. Concepts from the theory of dynamical systems that play a role in the theory are introduced. A state-space description is for the oscillator driven by a simple rhythmic signal is provided, and the results of a resonance analysis are given. This analysis yields an algorithm for simulating the behavior of the model, and it provides insight into how the entrainment model may provide the basis for a theory of meter perception. Chapter VIII describes tests of the model, showing how it is able to track beats in complex musical rhythms (collected in Chapter III) and how systems may be composed to model the perception of meter. Chapter IX describes the relationship between the two models, offers insights into how the models relate to other fields, and offers some closing thoughts.

CHAPTER II

SEQUENCE STRUCTURE AND TEMPORAL STRUCTURE IN MUSIC

This chapter presents a background of issues in music cognition that relate to the models that will be developed in subsequent chapters. First, perceptual grouping and recursive recoding, or chunking, are considered from the perspective of music cognition. The role of abstract structural relationships among events is also discussed. These issues are most relevant to the neural network model developed in Chapter IV. Next, the role of temporal structure in musical sequences is considered. This information will be important in the development of the entrainment model in Chapters VI and VII.

2.1 Rhythmic Grouping

The concept of information recoding, first introduced by Miller (1956), suggested that subjects presented with to-be-remembered sequences can reduce the amount of information to be retained by recoding, or chunking, subsets of more than one item into a single memory code. Researchers such as Estes (1972), Vitz and Todd (1969), and Garner and Gottwald (1968) argued that subjects assign codes to the subgroups of a sequence to reduce demands on memory, and these codes can be recalled and decoded on a later occasion to reconstruct the entire sequence. The principles proposed for grouping elements to produce codes were often perceptual; for example, Vitz and Todd (1969) suggested that runs of perceptually similar elements are cast into memory codes. For auditory patterns, perceptual grouping principles referred to rhythm.

The term *rhythm* refers to the sense of movement in time that characterizes the experience of music (Apel, 1972). One aspect of the perception of rhythm is the perception of the grouping of events (Cooper & Meyer, 1960; Lerdahl, & Jackendoff, 1983). *Phenomenal accent* is the physical patterning of events in the musical stream such that some seem stressed relative to others (Lerdahl, & Jackendoff, 1983). Perception of accent influences the perception of grouping. A great deal of research has gone into identifying the ways in which events can be accented, and in the effect of accent on the nature of the grouping percepts (for reviews see Fraisse, 1982 and Handel, 1989).

Phenomenal accent can be conferred upon the events of an auditory sequence by the manipulation of many possible physical variables including intensity, duration, and frequency. Different patterns of accentuation produce different grouping percepts. For example, if every second or third element of a sequence is accented by increasing its intensity, then the sequence is

perceived in groups of two or three, with the more intense event perceived as beginning the group (Fraisse, 1956). Similarly, if every second or third event is accented by increasing its duration, the sequence is perceived in groups of two or three, with the lengthened element ending the groups (Woodrow, 1951). If the inter-onset in interval (IOI) between every second or third element is lengthened, the sequence is again perceived in groups of two or three, however in this case the perception of accent depends on the length of the IOI (Povel, & Okkerman, 1981). Similarly, if an element of a sequence is accented by following a large pitch leap, it is perceived as beginning a group (Jones, 1981a).

In music these and other variables combine in complex ways to create grouping percepts. Lerdahl and Jackendoff (1983) have proposed a generative grammar to describe the perception of grouping in music. Their grammar is an attempt to describe general conditions for auditory pattern perception that have greater application than for music alone (Lerdahl, & Jackendoff, 1983). The theory describes a set of well-formedness rules that specify how a piece may be recursively subdivided into nested groups, or chunks. A set of preference rules attempts to summarize how physical variables influence the perception of grouping in music. Figure 2 gives an example of a rhythmic grouping for a simple melody. The lowest level groups correspond to the measures of musical notation, however, in more complex examples this is not necessarily the case.

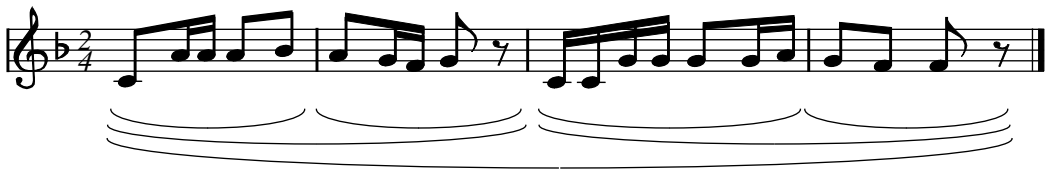


Figure 2: A rhythmic grouping for *Hush little baby.*

Recently, N. Todd (1994) has proposed a multiscale model of rhythmic grouping, based upon an analogy to visual edge detection. Using a direct temporal analog to the ∇^2 operator for spatial edge detection (Marr & Hildreth, 1980; Marr, 1982) Todd's system simultaneously carries out auditory edge detection at multiple time scales. The product of the analysis is a single hierarchical structure whose terminal elements are the onsets of individual events, capturing information about grouping structure and relative salience of events. The theory does not incorporate idiom-specific musical knowledge as does Lerdahl & Jackendoff (1983), however it does account for many known facts about accent and rhythmic grouping in a bottom-up fashion (N. Todd, 1994).

2.2 Patterns and Reductions

Grouping explains certain aspects of music perception and cognition. However, the recoding view (the idea the perceptual groups correspond to memory codes) has been criticized for its reliance on perceptual regularities. Grouping and recoding cannot explain listeners' abilities to predict

upcoming events in patterned sequences, or the ability to learn cultural and stylistic regularities. To explain such phenomena most theorists rely on structural descriptions of musical sequences. Depending on the musical dimension(s) under consideration the nature of the description will vary, but each relies on some abstract system of knowledge representing the underlying regularities of a particular musical style or culture. Through experience with a musical style, listeners are thought to internalize characteristic patterns of rhythm, melody, harmony, and so forth, which are used to integrate and organize musical sequences. Krumhansl argues that listeners abstract and internalize underlying regularities through experience with musical patterns. These cognitive representations give rise to expectations and affect the stability of memory (Krumhansl, 1979; Krumhansl, Bharucha, & Castellano, 1982). Jones (Jones, 1981a) argues that listeners abstract and store "ideal prototypes" of musical styles, that lead to musical expectations. Unexpected events in music create interest, but are more difficult to recall (Jones, Boltz & Kidd, 1982).

One set of theories applied to explain musical expectation and prediction are pattern-formation theories. Pattern-formation theories emphasize individuals' use of ordered vocabularies, or alphabets, and rules that apply to alphabetic properties. Several researchers (Jones, 1974; Restle, 1970; Simon & Kotovsky, 1963) have proposed that subjects abstract serial relations and, using rule-based transformations such as repeat, transpose, complement, and reflection, generate cognitive data structures that capture abstract relationships among events. The use of such transformations is thought to account for subjects' abilities to represent and predict unfolding serial patterns. Simon & Sumner (1968) extended serial pattern research to music, proposing that listening to music could be modeled as a process of pattern induction and sequence extrapolation, using alphabets and rule-based transformations such as *same* (repeat) and *next* (next element in the alphabet).

There are other types of musical phenomena that pattern-formation theories cannot explain, however. One is the extraction of invariant identification in musical variation (Large, Palmer, & Pollack, in press). This problem is interesting because the invariance of musical identity that characterizes the listener's experience is perceived across a wide range of differences in the surface content of the music (Dowling & Harwood, 1986; Lerdahl, & Jackendoff, 1983; Schenker, 1979; Serafine, Glassman, & Overbeeke, 1989; Sloboda, 1985). The problem of musical variation is best illustrated by an example. Consider the melodies of Figure 3. The melodies labeled A are the children's tunes *Hush little baby* (top), and *Mary had a little lamb* (bottom). The melodies labeled B are improvisations on these tunes, performed by pianists in an experiment described in Chapter III. Most listeners readily identify the B melodies as "variations" of the A melodies: listeners believe that the B melodies share an identity with the original melodies. However, one's listening to these examples or inspecting the musical notation will reveal that at the surface level these two sequences differ along many dimensions, including pitch content, melodic contour, and rhythm. Where is the similarity between these sequences?

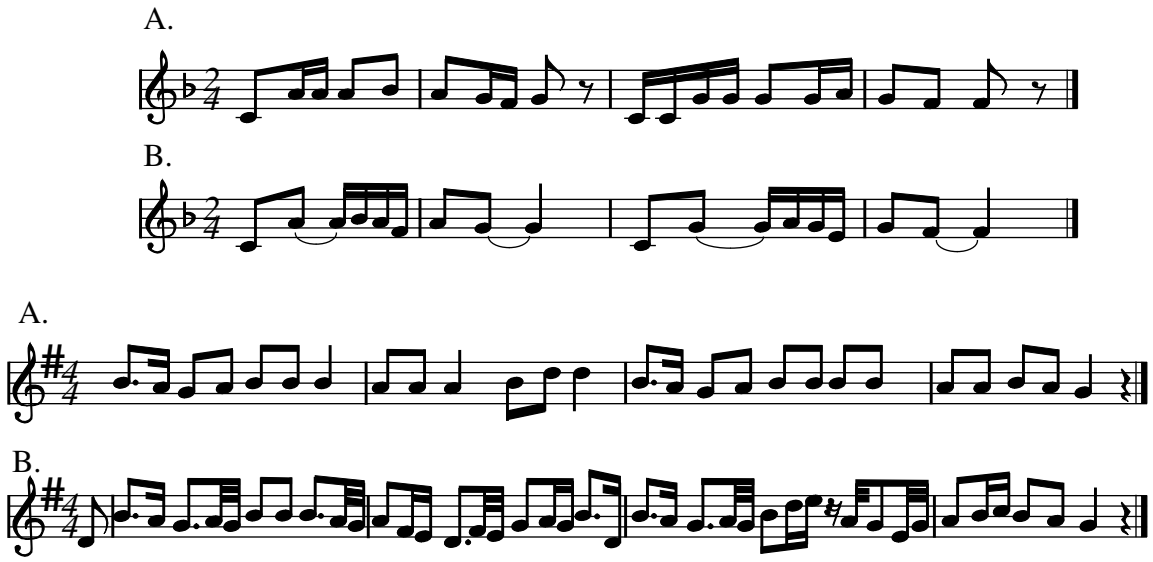


Figure 3: Melodies and variations. *Hush little baby* (top), and *Mary had a little lamb* (bottom), showing (A) subject melodies, and (B) improvised variations on the subject melodies.

One possibility is that as listeners produce internal representations for musical sequences, they implicitly evaluate the structural importance of events. Thus, certain events may be more important than others in determining the relationships that listeners hear between the melodies and variations of Figure 3. Evaluation of structural importance allows listeners to create reduced descriptions of musical sequences that retain the gist of the sequences while reducing demands on memory. *Reductionist theories* of music comprehension (Deutsch & Feroe, 1981; Lerdahl, & Jackendoff, 1983; Schenker, 1979) explain musical variation by positing a similarity of the underlying structures in related melodies.

One of the most comprehensive of the reductionist theories is Lerdahl and Jackendoff's "Generative Theory of Tonal Music" (Lerdahl, & Jackendoff, 1983). The theory takes as its goal the description of the musical intuitions of listeners experienced with Western tonal music. This is accomplished using a combination of music-theoretic analyses of which *time-span segmentation* and *time-span reduction* (illustrated in Figure 4) are the most relevant for current purposes. In Section 2.1 Lerdahl and Jackendoff's (1983) theory of grouping structure was considered as an approach to perceptual grouping. Viewed as part of a larger theoretical framework, however, the primary function of grouping structure is to identify temporal chunks that, in combination with an analysis of *metrical structure* (described below), exhaustively segments a musical sequence into rhythmic units called *time-spans*. The resulting *time-span segmentation* captures aspects of the piece's rhythmic structure, providing a hierarchically nested constituent structure description for the entire musical piece, shown by the brackets in Figure 4.

A time-span segmentation forms the input to a *time-span reduction* analysis, which organizes musical events into a structure that reflects a hierarchy of relative importance. Within each time-span a single most important event, called the head of the time-span, is identified. All

other events in the time-span are heard as subordinate to this event. The time-span reduction assigns relative importance to each event according to rules that consider melodic, harmonic, temporal, and structural factors. Thus, time-span reduction provides a unification of musical factors and predictions regarding which events listeners will perceive to be most important. Figure 4 shows a time-span reduction; the top musical staff shows the melody and the staves below show the heads for successively larger and larger time-spans. At each level, the less important event(s) of each time-span are eliminated, and a "skeleton" of the melody emerges. The tree above the top musical staff combines the information conveyed by the skeletal melodies with the information conveyed by the time-span segmentation. Its branching structure emphasizes structural relationships between levels of the reduction: that events of lesser importance are heard as elaborations of the more important events. The tree also identifies the structural ending of the musical passage, the cadence, shown as an ellipse "tying" together two branches of the tree, as shown in Figure 4.

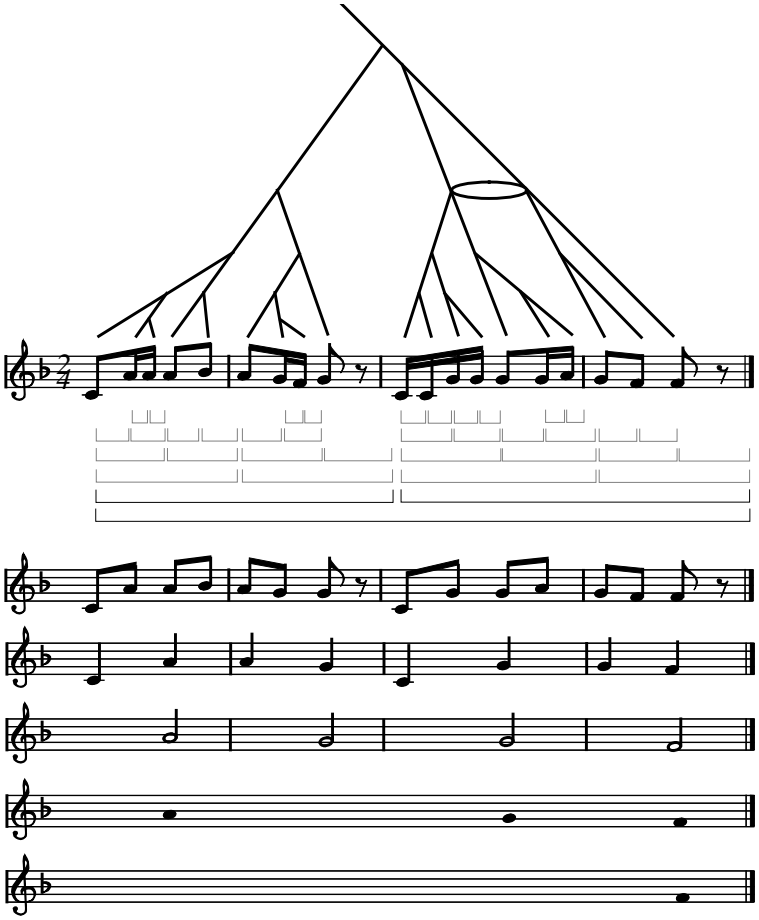


Figure 4: Analysis of *Hush little baby* following Lerdahl & Jackendoff (Lerdahl, & Jackendoff, 1983). The original melody is shown in musical notation and the brackets below mark the time-span segmentation. Solid brackets describe the contribution of grouping structure (Section 2.1); dotted brackets describe the contribution of metrical structure (Section 2.3). The tree above the notated melody is a time-span reduction. The lower staves show the dominant events for each level of the time-span segmentation.

Reductionist theories can be applied to explain the perception of musical variation. Figure 5 compares the theoretical reduction of the original melody *Hush little baby* with a reduction of an improvised variation on this melody (from Figure 3). At the third skeletal level, the two reductions are identical. Lerdahl and Jackendoff's (1983) theory can be applied to predict an intermediate level of mental representation at which structural similarities are captured.

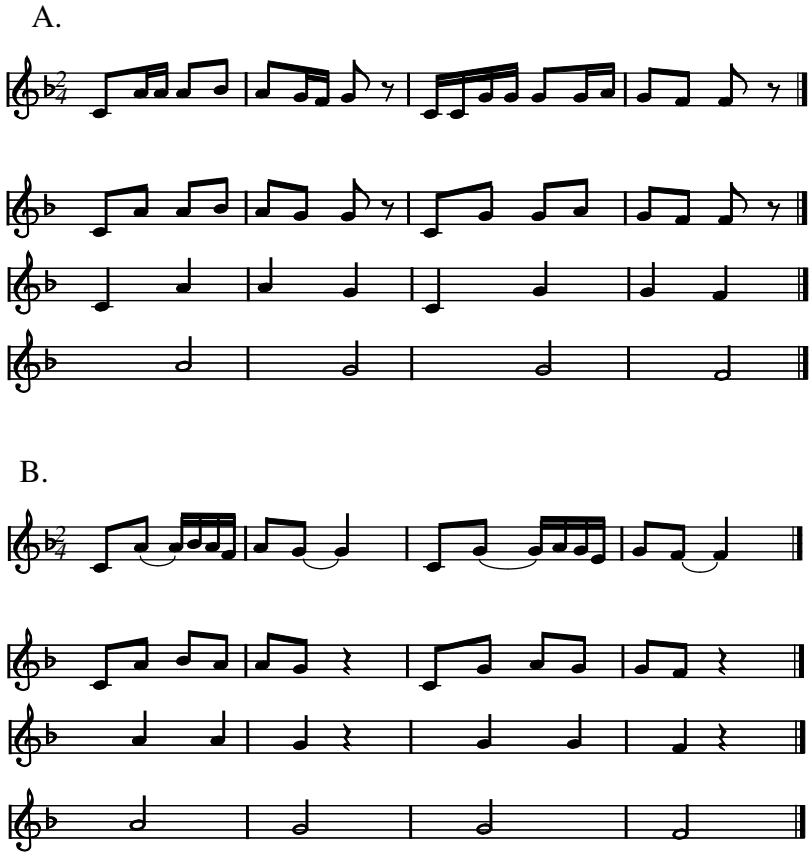


Figure 5: The first three levels of reduction for *Hush little baby* and its variation (from Figure 3). The reductions are identical at the third level.

2.3 Dynamic Attending, Temporal Expectancy, and the Entrainment Hypothesis

Hierarchical descriptions of musical structure are important; however, listeners experience music as temporal sequences. Understanding how musical sequences are comprehended *in time* is a central issue in music cognition. Meyer (1956) proposed that *expectation* is the key to understanding the human response to music. Through artful patterning of the acoustic environment, composers and performers evoke expectations in their listeners. They skillfully manipulate these expectations, satisfying some and frustrating others, to arouse both affective and intellectual responses. Meyer (Meyer, 1956) argued that this is the property of musical experience that enables artistic communication. Many theories, including structural theories described in the

previous section, have dealt with the issue of expectation, each exploring various types of expectancy in music perception and cognition. Simon and Sumner (Simon & Sumner, 1968) provide an analysis of music perception as a sequence extrapolation task, one in which the listener attempts to predict what patterns will follow based on analysis of the current pattern context. Narmour's implication-realization theory (Narmour, 1990) describes innate (culture-independent) expectancies that arise in response to the basic properties of individual melodic intervals and chains of melodic intervals, as well as expectancies that arise as a result of style-dependent knowledge. Lerdahl and Jackendoff's (1983) prolongational analysis describes the way music progresses from points of relative tension to points of relative repose.

Because time is the primary medium of musical communication, however, musical expectancy cannot be adequately characterized simply by considering *what* events a listener expects to occur. One must also consider *when* a listener expects events to occur (Jones, 1981b). In this regard, it useful to distinguish between sequential expectancy and temporal expectancy (Large & Kolen, in press). Sequential expectancy requires prediction of the sequential ordering of future events. Temporal expectancy requires anticipating when future events are likely to occur, utilizing knowledge of temporal structure.

Abstract knowledge of temporal structure has been shown to affect memory for sequence information in auditory sequences. In one study, memory for pitch sequences was found to be dependent on a perceived temporal frame. Pitch structures that coincided with temporal structures enhanced recall, while pitch structures that conflicted with temporal structures negatively affected recall (Deutsch, 1980). In a related finding, memory confusions of temporal patterns in a discrimination task were found to be consistent with a music-theoretic metrical structure hierarchy (Palmer & Krumhansl, 1990). Other studies have demonstrated similar memory constraints, by showing that the reproducibility of rhythms is affected by the patterns of phenomenal accentuation in the to-be-reproduced rhythm. The evidence suggests that sequences of events implying a metrical organization are easier to memorize and reproduce than sequences lacking such organization (Essens & Povel, 1985; Povel, & Essens, 1985). These and related findings are often cited as evidence that listeners represent and/or remember rhythms in terms of metrical structure hierarchies. Essens and Povel (1985) have hypothesized that in perceiving a temporal pattern, listeners induce an internal clock that is subsequently used as a measuring device to code the structure of a temporal pattern. Rhythmic sequences are encoded in memory with respect to this clock, so that patterns that correspond well with an induced clock (metrical patterns) can be represented using simpler memory codes, and are therefore easier to remember and reproduce.

Jones (1976; 1987) and Jones & Boltz (1989) offer an interpretation known as *dynamic attending*. They argue that the organization of perception, attention, and memory is inherently rhythmical. Music (and other rhythmic stimuli) *entrains* listeners' perceptual rhythms, and these rhythms embody *expectancies* for when in time future events are likely to occur. Expectancies in turn guide *anticipatory pulses of attention* that facilitate perception of events that occur at expected points in time. Dynamic attending is a theory of temporal expectancy that can be applied to the perception of music, among other things.

One source of evidence for dynamic attending stems from studies that directly test listener attention rather than listener memory. These studies show that temporal pattern structure constrains the ability of subjects to attend to melodic sequences. For example, regularity of phenomenal accent placement has been shown to affect listeners' abilities to judge the temporal order of tones in a sequence (Jones, Kidd, Wetzel, 1981). Listeners are also better able to identify pitch changes in sequences when these changes occur at points of strong metrical accent (Jones, Boltz & Kidd, 1982). Additional evidence suggests that listeners' implicit knowledge of musical meter (beyond immediate sensory context) contributes to the perception of temporal sequences. Listeners' goodness-of-fit judgements for events presented in metrical contexts were shown to be consistent with multi-leveled metrical structure hierarchies (Palmer & Krumhansl, 1990).

Dynamic attending is a complex theory. Part of this theory refers to the synchronization of perceptual processes to temporally structured event sequences. I shall call this the *entrainment hypothesis*. A source of evidence supporting perceptual entrainment comes from psychophysical studies of time perception. The temporal structure of auditory patterns affects humans' abilities to perceive time. For inter-onset durations corresponding roughly to musical time scales, it can be shown that the ability to detect differences in temporal intervals approximately obeys Weber's law (Getty, 1975; Halpern & Darwin, 1982). That is, when subjects are asked to compare two intervals, the accuracy of their time discrimination judgement is related to the base length of the interval they are asked to judge. Adherence to Weber's law breaks down under certain circumstances, however. Temporal difference judgements improve as the number of reference intervals increases (Schulze, 1989; Drake & Botte, 1993). It has also been shown that sensitivity to time changes in sequences is best for metrically regular sequences (Yee, Holleran & Jones, in press), and that sensitivity to tempo changes degrades with the regularity of the stimulus (Drake & Botte, 1993). Some researchers have suggested that these results indicate perceptual synchronization of the listener to a perceived beat (e.g. Schulze, 1989; Yee, Holleran & Jones, in press), supporting the entrainment hypothesis.

Theories of temporal structure gain complexity when applied to the musical case. The music-theoretic notions relevant to the current discussion are *beat* and *meter. Beat* refers to one of a series of perceived pulses marking (subjectively or perceptually) equal units in the temporal continuum. Beat perception is established and supported by musical events, however, once a sense of beat has been established, it continues in the mind of the listener even after the event train has ceased (Cooper & Meyer, 1960). The term *tempo* refers to the frequency (beats per unit time) at which beats occur. The reciprocal measure, *beat period*, refers to the span of time between consecutive beats. According to the entrainment hypothesis, beat perception is a form of temporal expectancy – the perception of beats corresponds to the expectation that events will occur at roughly equal intervals – thus, a beat *is* the expectation of an event. This simple form of temporal expectancy enables a more complex form of temporal expectancy, the perception of *meter.*

Simply defined, *meter* is the number of beats between (more or less) regularly recurring phenomenal accents (Apel, 1972; Cooper & Meyer, 1960). Meter can be described as two or more periodicities in a sequence of events, corresponding to multiple levels of beats perceived on different time scales (Cooper & Meyer, 1960; Lerdahl, & Jackendoff, 1983; Yeston, 1976). Lerdahl

and Jackendoff (1983) have proposed a construct that describes the temporal organization of a piece at all relevant metrical levels, called a *metrical structure*. The metrical structure of a piece can be transcribed as a grid (Figure 6). According to this notation, each horizontal row of dots represents a level of beats, and the relative spacing and alignment among dots of adjacent levels captures the relationship between the periods and phases of adjacent levels of beats. Metrical structure grids describe an important component of rhythmic experience: the perception of regularly recurring strong and weak beats called *metrical accent* (Lerdahl, & Jackendoff, 1983). Points of metrical accent are captured, using the grid, as temporal locations where the beats of many levels coincide. Points where many beats coincide are called *strong beats*; points where few beats coincide are called *weak beats*.

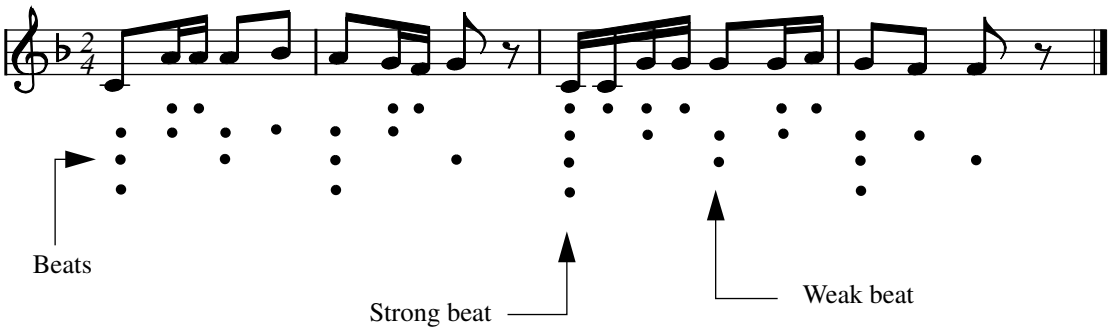


Figure 6: A metrical structure for *Hush little baby.*

Theories of meter perception describe an important aspect of rhythm perception. A rhythm, with its pattern of phenomenal accent, is thought to function as a perceptual "input" from which the listener may extrapolate a regular metrical pattern (Lerdahl, & Jackendoff, 1983). Lerdahl and Jackendoff (1983) have proposed a generative theory of meter perception expressed as two sets of rules. A set of *well-formedness rules* describes legal metrical structures. These rules restrict metrical structures to strictly nested hierarchies with beat-period ratios of either 2:1 or 3:1. Next, a set of *preference rules* describes which legal metrical structure an experienced listener would perceive for a given rhythmic pattern. These rules are concerned mainly with the placement of strong beats, as determined by the alignment of beats at adjacent levels in the metrical structure grid.

The restriction to strictly nested beat periods has certain advantages within the context of Lerdahl and Jackendoff's (Lerdahl, & Jackendoff, 1983) theory. For one thing, it allows metrical structure analysis to act in concert with grouping structure to segment a piece into nested time spans (see Section 2.2). However, this characterization of metrical structure limits the scope of the theory (Lerdahl, & Jackendoff, 1983). Much non-Western music, contemporary Western Art music, Jazz, and popular music makes use of *dissonant* rhythmic structures (Yeston, 1976), known as *polyrhythms* (Figure 7). A polyrhythmic relationship between two levels of beats is a relationship of beat-periods such that N beats at one level occupy the same amount of time as M

beats at the next level. *Rational* ratios N:M, such that the integers N and M are relatively prime (3:2, 4:3, 5:4, and so forth), characterize polyrhythmic ratios. Hierarchical nestings do not adequately capture polyrhythmic structures. It is more general to think of metrical structures as composed of layers, or *strata*, of beats at different time scales (Yeston, 1976).

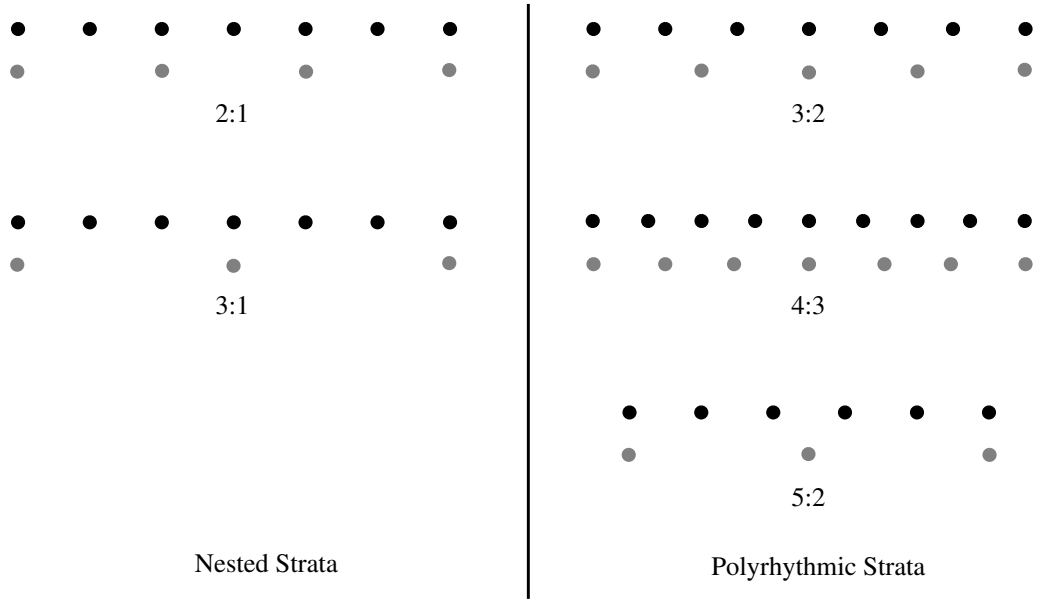


Figure 7: Metric strata. Simple ratios imply grouping. Polyrhythmic ratios do not imply grouping.

A limitation of current theories of meter is that they fall short of adequately explaining perception. Theories of metrical structure, as discussed above, apply to musical time as notated. It is well established, however, that musicians never perform rhythms in a regular, or mechanical, fashion. Instead, performers produce sound patterns that reveal both intentional and unintentional timing variability (Clarke, 1985; Drake & Palmer, 1993; Palmer, 1988; Shaffer, Clarke & N. Todd, 1985; Sloboda, 1983; N. Todd, 1985). Current theories of metrical structure do not explain how listeners are able to perceive meter in rhythms that performers actually play (unless the performer is a computer).

In summary, theories of metrical structure attempt to describe the perceived temporal organization of rhythmic patterns. A metrical structure is composed of layers, or strata, of beats that align with the onset of musical events. Theories of metrical structure address issues related to the beat period ratio and the relative alignment between adjacent levels of beats. Theories that require the layering of beats to describe a strictly nested hierarchy, however, are limited in scope. To include the polyrhythmic structures common in many forms of music, more complex relationships between adjacent levels must be allowed. Finally, theories that do not deal with the issue of timing variability in music performance, stop short of explaining the perception of metrical structure.

CHAPTER III

A STUDY OF MUSIC PERFORMANCE AND IMPROVISATION

This chapter describes an empirical study of the performance and improvisation of melodies by skilled pianists. The data is analyzed for two purposes. First, the improvisations are analyzed to determine the nature of structural relationships in performers' mental representations of three melodies. A measure of relative importance of events for each melody is extracted based on the improvisations. These measures are compared to the predictions of a reductionist theory (Lerdahl, & Jackendoff, 1983). In Chapter IV, this data will be used to test a neural network that produces structured descriptions for musical sequences. Second, performances and improvisations are analyzed to determine the nature of the timing in these skilled performances. Temporal structure is first analyzed using the methods of Fourier analysis and auto-correlation. Next, systematic timing deviation, or *rubato*, is analyzed by comparing performed event durations to *ideal* durations, determined from scores of the melodies and transcriptions of the improvisations. In Chapter VIII this data will be used to test a model of the perception of temporal structure in music.

3.1 Structural Relationships Among Sequence Events

The first analysis investigates the relative importance of events in performers' mental representations of three melodies. Empirical evidence supporting reductionist descriptions of structural relationships among events has emerged in the literature. Previous studies have dealt primarily with perceptual phenomena (Palmer & Krumhansl, 1987a; 1987b; Serafine, Glassman, & Overbeeke, 1989). The reductionist hypothesis also leads to predictions concerning music performance. For example, in musical traditions that employ improvisation, performers may identify the gist of a theme in terms of its structurally important events and use techniques of variation to create coherent improvisations on that theme (Johnson-Laird, 1991; Lerdahl, & Jackendoff, 1983; Pressing, 1988). Therefore, it should be possible to identify the events of greater and lesser importance in a melody by collecting improvisations on that melody and measuring the events that are retained across improvisations. This rationale is used to identify structurally important events by asking performers to improvise variations on a melody; the variations are examined for events altered or retained from the original melody.

Several methods have been employed to elicit the structure of listeners' mental representations for musical sequences. In one study (Palmer & Krumhansl, 1987a; 1987b) subjects were asked to listen to excerpts from a musical passage and rate how "good or complete" a phrase

each excerpt formed. The rating was taken as a measure of the relative importance for the final event in each musical excerpt. Listeners' judgements of phrase completion at various points in a musical passage correlated well with predictions of each events' relative importance from time-span reductions (Palmer & Krumhansl, 1987a; 1987b; Lerdahl, & Jackendoff, 1983). The nature of the musical task, however, was somewhat unnatural, because music is usually not presented in fragments. Additionally, the application of this paradigm to longer musical works is problematic. In another study listeners were asked to judge the similarity between related melodies (Serafine, Glassman, & Overbeeke, 1989). Although this paradigm does not provide measures of importance for individual musical events, it does allow the assessment of reductionist claims within an ecologically valid task. The similarity judgements among melodies corresponded to the degree of relatedness predicted by a reductionist theory (Schenker, 1979), even when radical surface differences existed (such as in the musical harmony). This agreement increased with repeated hearings, indicating a significant role of learning in determining the structure of listeners' mental representations (Serafine, Glassman, & Overbeeke, 1989).

The experiment reported here is based on a paradigm described earlier (Large, Palmer, & Pollack, 1991). In this paradigm, musicians are presented with notated melodies and are asked to improvise (create and perform) simple variations on them. Improvisation in Western tonal music commonly requires a performer to identify some framework of melodic and harmonic events, and apply procedures to create elaborations and variants on them (Johnson-Laird, 1991; Steedman, 1982; also, see Pressing, 1988 for a review of improvisational models). Thus, improvisation of variations allows musicians freedom to determine which if any musical events should be retained from the original melody. This paradigm addresses the reductionist account by measuring musicians' intuitions about a particular melody within the context of a familiar task. This paradigm has an additional advantage in that it allows for the collection of individual ratings of importance for each event. Musical events viewed as structurally important should tend to be retained in improvised variations. Events viewed as less important (i.e., events that function as elaborations of important events) should be more likely to be replaced with different elaborations.

The relative importance of each pitch event in the original melody is measured by counting the number of times it was retained in the same relative temporal location across improvisations. Although this is a coarse measure of improvisation, it allows one to generalize across many aspects specific to music performance (including dynamics, phrasing, rubato, pedaling, etc.) and improvisation (including motific development, stylistic elaboration, etc.), and concentrate instead on those factors that reflect reductionist considerations.

The primary objective of this study was to extend findings (Large, Palmer, & Pollack, 1991) that suggested that a musician's improvisations on a tune indicated an underlying reduced representation of the melody. According to the application of the time-span reduction hypothesis to improvisation, more important events (those retained across multiple levels of the time-span reduction) should be more likely than unimportant events to be retained in variations on a melody. Therefore, the number of individual pitch events retained in the musicians' improvisations should correspond to the theoretical predictions of structural reductions.

3.2 Temporal Structure in Music Performance and Improvisation

The second analysis investigates the nature of timing in the performance of three melodies and in a set of improvised variations on these melodies. In Chapter VIII this data will be used to test a model based on the *entrainment hypothesis* introduced in Chapter II. A great deal of evidence supporting the entrainment hypothesis already exists in the empirical literature. Some studies have dealt with primarily perceptual phenomena (e.g. Jones, Kidd, Wetzel, 1981; Schulze, 1989; Drake & Botte, 1993; Yee, Holleran & Jones, in press), and others have investigated the ability of listeners to synchronize motor behavior with auditory rhythms (e.g. Fraisse, 1956; Povel, & Essens, 1985; Essens & Povel, 1985). However, most studies have investigated responses to temporal patterns that are much simpler than those found in musical performance. The synchronization hypothesis also leads to predictions concerning the perception of complex temporal structures in skilled music performance. Performers deviate from mechanical regularity, yet timing deviations rarely pose difficulty for human listeners. It is possible to quantify the amount of deviation from timing regularity by comparing the actual timing of performances and improvisations with ideal temporal relationships, as found in scores or transcriptions of musical performances (e.g. Bengtsson & Gabrielsson, 1983, Palmer, 1988). This rationale is used to assess the amount of perceptual flexibility that listeners *must* possess to successfully cope with temporal patterns of the complexity of performed musical rhythms.

Performance timing has proved to be an especially fertile area of study in music cognition, and the basic findings are consistent. Rhythms performed by skilled musicians show deviations from timing regularity (as prescribed by the musical score) that are systematically related to the musical intentions of performers (Drake & Palmer, 1993; Clarke, 1985; Palmer, 1988; 1989; Shaffer, Clarke & N. Todd, 1985; Sloboda, 1983; N. Todd, 1985). It is assumed that listeners respond to these perceptual cues and comprehend the intentions of performers, so deviation from *ideal* timing in musical performance communicates musical information. The assumption that listeners respond to timing deviation entails that listeners are somehow able to recover the idealized timing of the musical score. Precisely how listeners are able to do this remains a subject of debate. Perceptual studies have investigated this ability, for example studies of time perception have investigated listeners' ability to detect small individual deviations from timing regularity in carefully controlled temporal sequences (see Section 2.3). These studies offer a valuable source of information regarding time perception. Applicability to musical sequences is limited, however, because of the rigorous controls employed in stimulus construction.

Music performance and improvisation provide a potentially rich domain for the study of time perception in complex sequences. Because performed musical sequences contain systematic timing deviations, musical performance provides a complex and ecologically valid source of data on time perception. The analyses reported here are based on the observation that although actual inter-onset durations measured in skilled musical performance are more-or-less out of time (as prescribed by the musical notation), listeners perceive inter-onset durations in terms of *ideal* duration categories corresponding to the quarter-notes, half-notes, measures, and so forth, of musical notation (Clarke, 1989; Longuet-Higgins & Lee, 1982). Thus, the difference between

notated and observed timing relationships provides an accurate measure of the temporal deviations with which listeners cope in the perception of complex musical passages.

The temporal structure of performances and improvisations is measured in three ways. First, skilled analysts transcribe musical improvisations in standard musical notation and agree upon the transcriptions. This provides a baseline measure of perceived temporal structure (this step was not necessary for the performed melodies). Next, the techniques of Fourier analysis and auto-correlation are used to compare the *ideal* temporal relationships described by the notation with the *actual* timing observed in the performances. An advantage of such methods is that they do not rely upon postulated mental processes, rather they examine "physical properties" of notated scores and performances (Brown, 1993). Finally the amount of rubato in the performances was characterized as average deviation from *ideal* timing, by comparing performed durations with scores and transcriptions.

One thing this data will not tell us is *how* listeners cope with timing deviation. The goal of this analysis is to characterize a set of data that will be used to test the model of beat perception described in Chapters VI and VII. In these chapters, beat perception is modeled as a dynamical system in which an oscillator is non-linearly coupled to a rhythmic driving stimulus. Rather than directly observing the trajectories of the dynamical system (as would be done in perceptual studies) assumptions about what the trajectory *should* be will allow use of this rich source of data to evaluate the trajectories observed in the computer model.

3.2.1 Method

3.2.1.1 Subjects

Six skilled pianists from the Columbus, Ohio community participated in the experiment. The pianists had a mean of 17 years (range of 12 to 30 years) of private instruction, and a mean of 24 years (range of 15 to 32 years) of playing experience. All of the pianists were comfortable with sight-reading and improvising. All were familiar with the pieces used in this study.

3.2.1.2 Materials

Three children's melodies (*Mary had a little lamb*, *Baa baa black sheep*, and *Hush little baby*) were chosen as improvisational material that would be familiar (well-learned) for most listeners of Western tonal music, to insure a well-established notion of relative importance for each event and to avoid learning effects. Additionally, these pieces were fairly unambiguous with regard to their time-span reductions.

3.2.1.3 Apparatus

Pianists performed on a computer-monitored Yamaha Disklavier acoustic upright piano. Optical sensors and solenoids in the piano allowed precise recording and playback without affecting the

touch or sound of the acoustic instrument. The pitch, timing, and hammer velocity values (correlated with intensity) for each note event were recorded and analyzed on a computer.

3.2.1.4 Procedure

The following procedure was repeated for each piece. Pianists performed and recorded the melody, as presented in musical notation, five times. These initial recordings allowed each pianist to become acquainted with the improvisational material. With the musical notation remaining in place, the pianists were then asked to play five "simple" improvisations. The pianists were also asked to play five "more complex" improvisations, which are not discussed here. All performances were of a single-line melody only; pianists were instructed not to play harmonic accompaniment. All pianists indicated familiarity with all of the musical pieces.

3.3 Analysis #1: Mental Representation of Melodies

3.3.1 Coding Improvisations

Each improvisation was coded in terms of the number of events retained from the original melody, to develop a measure of relative importance for each event. The following procedure applied to the coding of each improvisation. First, the improvisation was transcribed by two musically trained listeners, who agreed on the transcriptions. Next, sections of the improvisation were matched to sections of the original. For most improvisations this was straightforward; for two of the improvisations, sections that repeated in the original melody (*Baa baa black sheep*) were rendered only once in the improvisation, and these were doubled for purposes of analysis. Finally, individual events of the improvisation were placed into correspondence with the original. If only the pitch contents and rhythm changed (meter and mode remained the same), as in most of the improvisations, this process was straightforward: events were placed into correspondence by metrical position. For mode change (for example, the flatted third is substituted for the major third in a major to minor mode shift), substitutions were counted as altered events. For meter change, metrical structures were aligned according to the onsets of each measure and half-measure, and events were then placed into correspondence by temporal location. Those events whose pitch class was retained in the correspondence between original melody and variation were coded as "hits" and received a score of 1; those events whose pitch class was altered (or for whom no event corresponded in the improvisation) were coded as "misses" and received a score of 0. For example, if a quarter note "C" were replaced with four sixteenth notes "C-B-C-B" beginning at the same metrical location, the "C" would be coded as a hit. If, however, the "C" had been replaced with "B-C-B-C", the "C" would be coded as a miss. Thus, only deletions and substitutions of events in the original melody affected the number of hits.

The number of hits for each pitch event in the original melody was summed across the five improvisations for each performer. To rule out the possibility that events in the original melody were altered at random, or that performers simply added events to create improvisations, an analysis of variance (ANOVA) on performers' mean number of retained events by event location

was conducted for each melody. Each of the three ANOVAs indicated a significant effect of event location (Melody 1: $F(25,125) = 4.02$, $p < 0.01$; Melody 2: $F(52, 260) = 6.64$, $p< 0.01$; Melody 3: $F(18, 90) = 7.76$, $p < 0.01$). Thus, performers were more likely to retain some melodic events than others across improvisations. The factors influencing the number of retained events at each location were further investigated in the following analyses.

3.3.2 Comparison with Theoretical Predictions

Theoretical reductions (Lerdahl, & Jackendoff, 1983; see Section 2.2 on page 10) can be quantified, as shown in Figure 8; the numbers below the time-span segmentation correspond to the relative importance of each event described by the time-span reduction analysis. Each number is a count of the number of branch points passed in traversing the tree from the root to the branch that projects in a straight line to the event, inclusive. For instance, to calculate importance for the first note of the melody, count 1 for the root, 1 for a left turn, 1 for a branch point passed, and 1 for a second left turn. This final branch projects in a straight line to the event, and so counting stops. The branch leading to the first event of a cadence is not counted as a branch point because it is considered structurally as "part of" the final event (Lerdahl, & Jackendoff, 1983). For example, to calculate importance for the first note of measure three, count 1 for the root, 1 for a branch point passed, 0 for a left turn (because this branch is tied), and 1 for a second left turn. According to this strategy, the smaller the number, the more important is the corresponding event. Metrical accents also make predictions of relative importance based on event location. These predictions can be quantified by numbering the levels of beats that correspond to metrical predictions (the numbers next to the metrical structure grid in Figure 8). The two measures are usually correlated because time-span reduction is partially based on metrical accent, but the time-span reduction adds information beyond metrical structure. Both quantifications of relative importance and metrical accents (computed similarly to Palmer & Krumhansl, 1987a; 1987b), will be compared with the measures from improvisational music performance.

Correlations between improvisation measures and both sets of theoretical predictions for each melody are summarized in Table 1. First, the correlation between the number of pitch events retained and the quantified metrical accent predictions for each event location were significant for each melody ($p < 0.05$). Improvisation measures were next compared with predictions from the time-span reduction analysis for each melody, obtained by quantifying the number of branch points passed in the tree, from root to terminal branch, as shown in Figure 8. Correlations between the number of pitch events retained and the quantified time-span reduction predictions were also significant for each melody ($p < 0.05$).

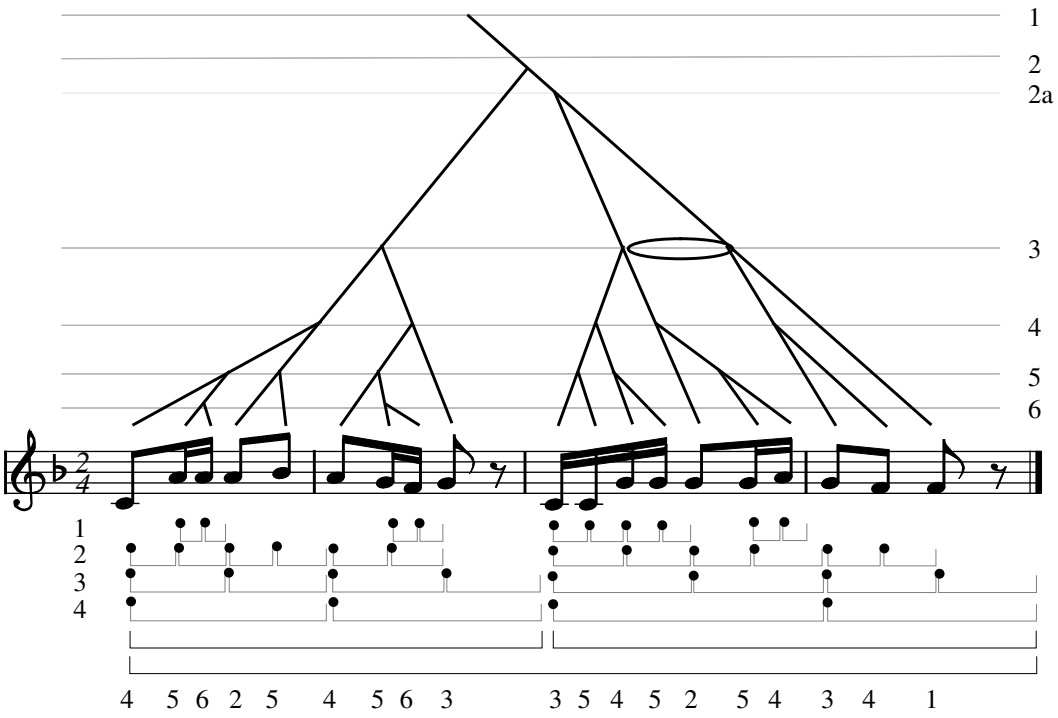


Figure 8: Analysis of *Hush little baby* showing metrical structure, time-span segmentation, and time-span reduction. The quantifications of relative importance for each event are shown below the segmentation.

Table 1: Squared correlation coefficients for theoretical predictions and improvisation-based measures.

	Melody 1: (*Mary*)	Melody 2: (*Baa*)	Melody 3: (*Hush*)
Metrical Accent Predictions	.63*	.80*	.78*
Time-Span Predictions	.76*	.79*	.67*
Semi-Partial (metrical accent removed)	.42*	.30*	.21

*-$p < 0.05$

To insure the predictive power of the time-span reduction beyond metrical accent, the improvisation measures were correlated with time-span reduction predictions after the effects of metrical accent were partialled out. These semi-partial correlations, also shown in Table 1, were significant ($p < 0.05$) for melodies 1 and 2, indicating that time-span reduction did contribute information beyond metrical accent. The semi-partial correlation was not significant for melody 3 ($p = 0.37$), indicating that in this case correlation of improvisation measures with the time-span reduction analysis was largely due to the effects of metrical accent.

3.4 Discussion

Musicians' improvisations of variations on simple melodies provided strong support for the reductionist hypothesis. Performers tended to retain certain events in each melody, and used improvisational techniques to create variations around those retained events. In addition, the music performances agreed with reductionist predictions of which events were relatively important in these simple melodies. Furthermore, the findings for two of the three melodies indicate that musical factors specific to time-span reductions played an important role in musicians' improvisation of variations.

The relatively high contribution of metrical structure to the improvisations based on the third melody (*Hush little baby*) may indicate a qualitative difference between the performers' intuitions and the theoretical predictions for this piece. For example, the improvisations often retained the first event of measure 1, an indicator of its relative importance, disagreeing with the theoretical weighting of this event. This may be due to the salience of the large initial pitch interval, or it may be a general primacy effect (making the first few events more likely to be retained despite reductionist considerations). The performances also disagreed with the predictions at the structural ending; all events in measure 4 were retained relatively often. Alternatively, this could be accounted for as a recency effect.

These discrepancies emphasize the difficulty of providing a relative weighting for a set of rules that determine the reductionist structure of mental representations. For example, the particular order in which a subset of rules is applied can lead to different weightings of constituents. However, the improvised performances do show general agreement with the theoretical predictions of time-span reduction. This is the first demonstration, to my knowledge, that the musical factors incorporated in the reductionist theory (Lerdahl, & Jackendoff, 1983) can account for the structure of performers' mental representations for musical improvisations.

3.5 Analysis #2: Performance and Improvisation Timing

3.5.1 Timing Analyses

This analysis considered both performances and improvisations for the first two pianists. First, *ideal* performances were developed in which each inter-onset duration had the precise relative duration prescribed by the musical score. For the performances this was straightforward, since the

pianists performed from musical notation. For the improvisations *ideal* performances were derived from the transcriptions prepared in the previous analysis. Absolute durations for the ideal performances were calculated based on one quarter note equals 480 ms (2.08 Hz).

A Fourier analysis was performed on the *ideal* performances. Fourier analysis is a spectral analysis that yields a frequency representation of an input signal. For this analysis, the input signal, $s(t)$, to the discrete Fourier transform (DFT) was composed of *unit samples*, $s(t) = 1$, at the onset of an event in the *ideal* performances, and $s(t) = 0$ at all other times. Thus, onset impulses did not carry amplitude information. Representative analyses are shown in Figures 9 and 10. Figure 9 shows the analysis of the *ideal* performance of *Mary had a little lamb* and Figure 10 shows an analysis of an *ideal* (as transcribed) improvisation on *Hush little baby.* The results of the Fourier analyses are shown in panels (C). The analyses both show characteristic patterns, with peaks at locations that reflect the metrical structures of the respective melodies.

Next, the auto-correlation function of the signal was calculated for each *ideal* performance, shown in panels (D) of Figures 9 and 10. Auto-correlation turns up peaks that reflect the metrical structure of the *ideal* performances because in Western melodies more events occur at strong metrical locations (Brown, 1993; Palmer & Krumhansl, 1990). The DFT of the auto-correlation function was then computed. The results are shown in panels (E) of Figures 9 and 10. Amplitude information added by auto-correlation improves the results of Fourier analysis. The output of the DFT is smoothed, and the peaks that correspond to the metrical structure are enhanced.

Analyses were repeated using input signals derived from the *actual* performances and improvisations to compare their temporal structure with that of their *ideal* counterparts. Figure 11 shows the analysis of an actual performance of *Mary had a little lamb* and Figure 12 shows an analysis of the actual improvisation on *Hush little baby*, from which the transcription of Figure 10 was prepared. As determined from the analyses described in the next section, mean deviation from average tempo for these two melodies was 5% and 15%, respectively. As the figures show, energy is added to the Fourier spectra across a wide range of frequencies. This has the effect of blurring characteristic patterns, making it difficult to match the frequency components of the *actual* performances to the spectral peaks generated by the *ideal* timing patterns.

The auto-correlation turned up few peaks corresponding to strong metrical locations for the *actual* performances, although the DFT of the auto-correlation function is somewhat improved compared to the DFT of the raw signal, indicating that auto-correlation is adding some information. Overall, however, the output is still rather difficult to interpret. One reason for this difficulty can be seen by comparing Figure 9, panel (D) with Figure 11, panel (D). The modal inter-onset duration in this piece is the eighth note (240 ms in *ideal* performances) corresponding to the peak at 4.17 Hz in Figure 9. The average duration for an eighth note in the actual performance was 299 ms, a frequency of 3.34 Hz. Around this point in Figure 11, however, there are two peaks. The existence of multiple spectral peaks is due to the fact that the frequency components change throughout this performance. This characteristic of musical performance, systematic frequency change at metrical time-scales, is called *rubato*.

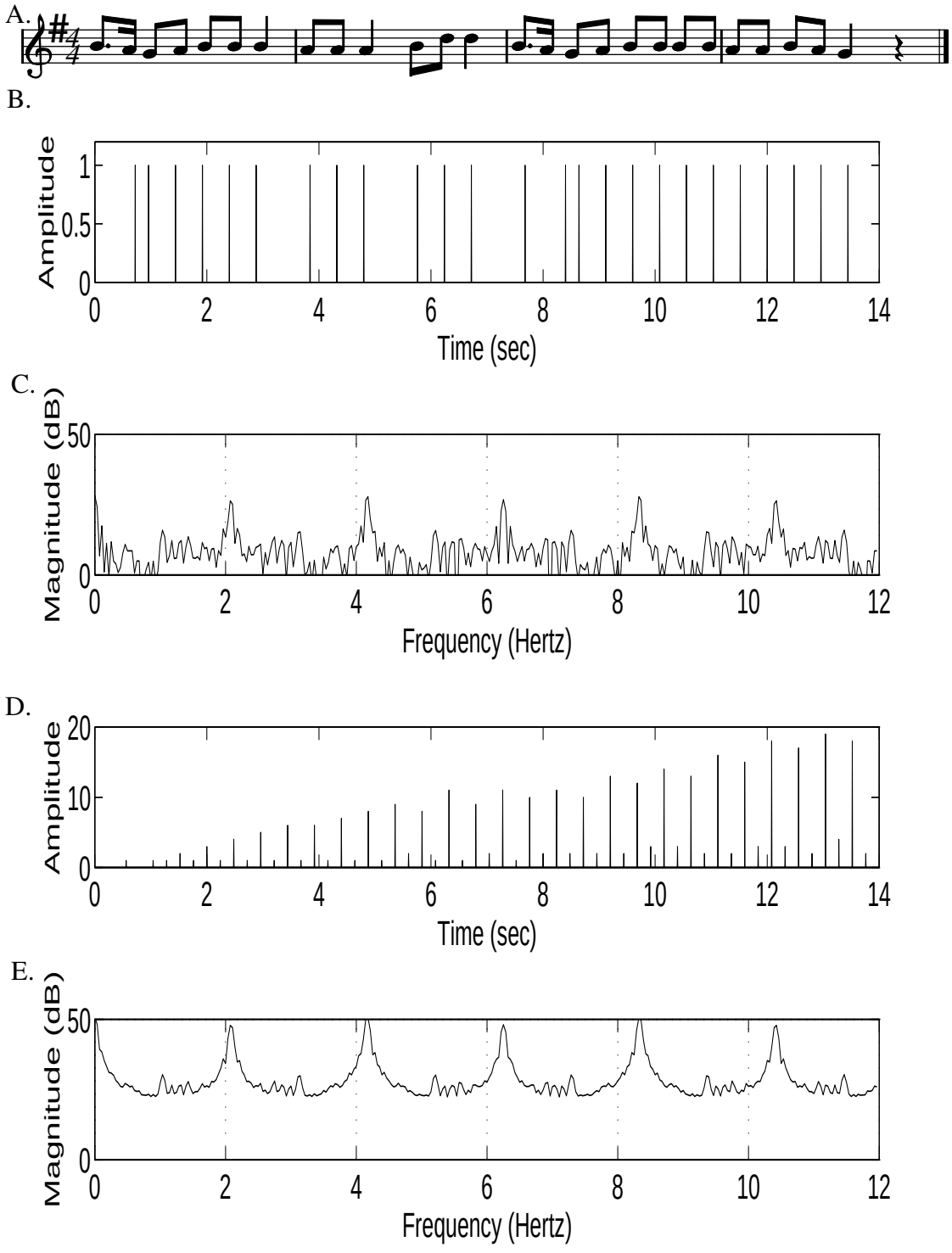


Figure 9: *Mary had a little lamb*: (A) musical notation, (B) onsets times prescribed by score, (C) discrete Fourier transform of B, (D) auto-correlation function of B, (E) discrete Fourier transform of D.

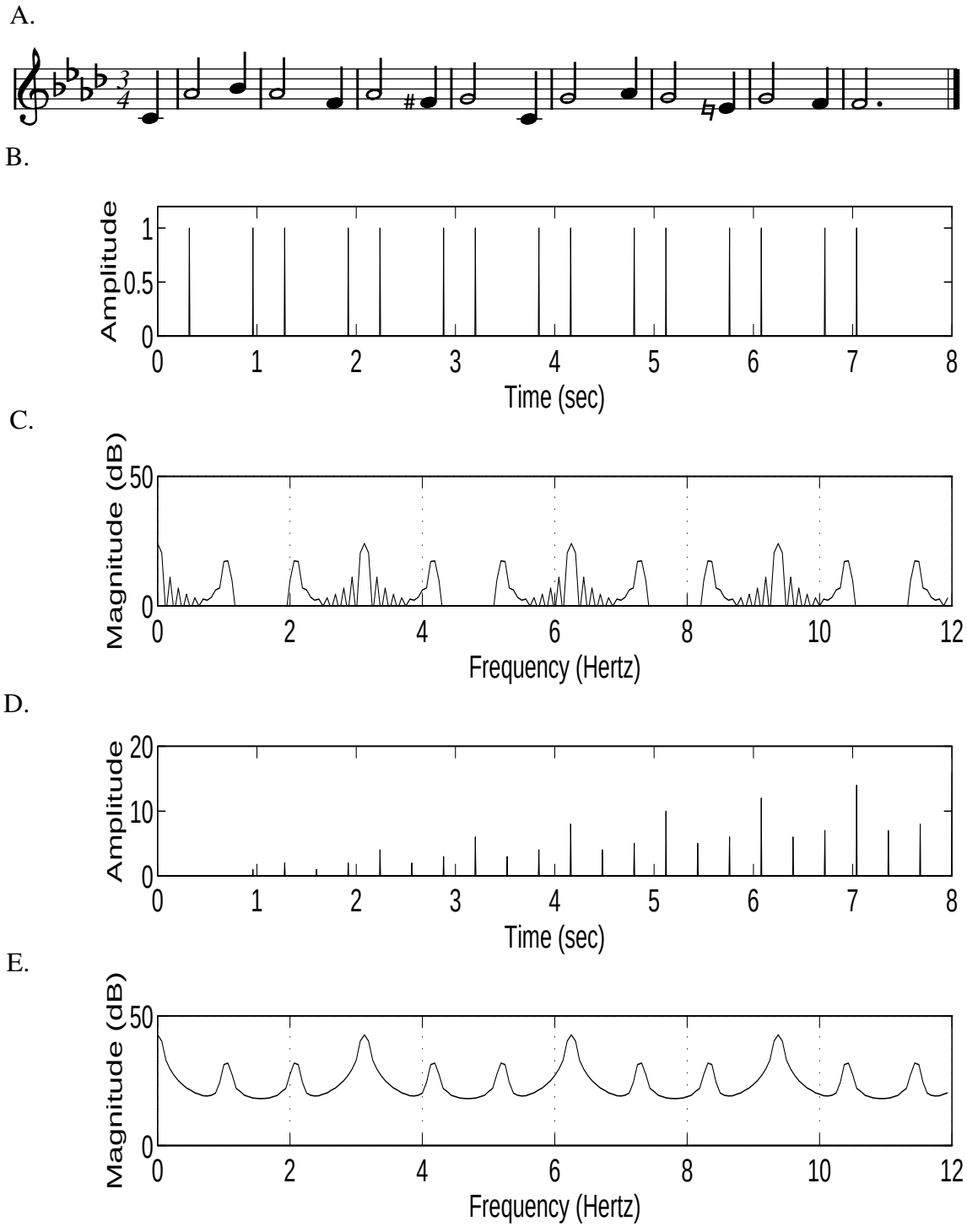


Figure 10: Transcription of an improvisation on *Hush little baby*: (A) musical notation, (B) onsets time prescribed by score, (C) discrete Fourier transform of B, (D) auto-correlation function of B, (E) discrete Fourier transform of D.

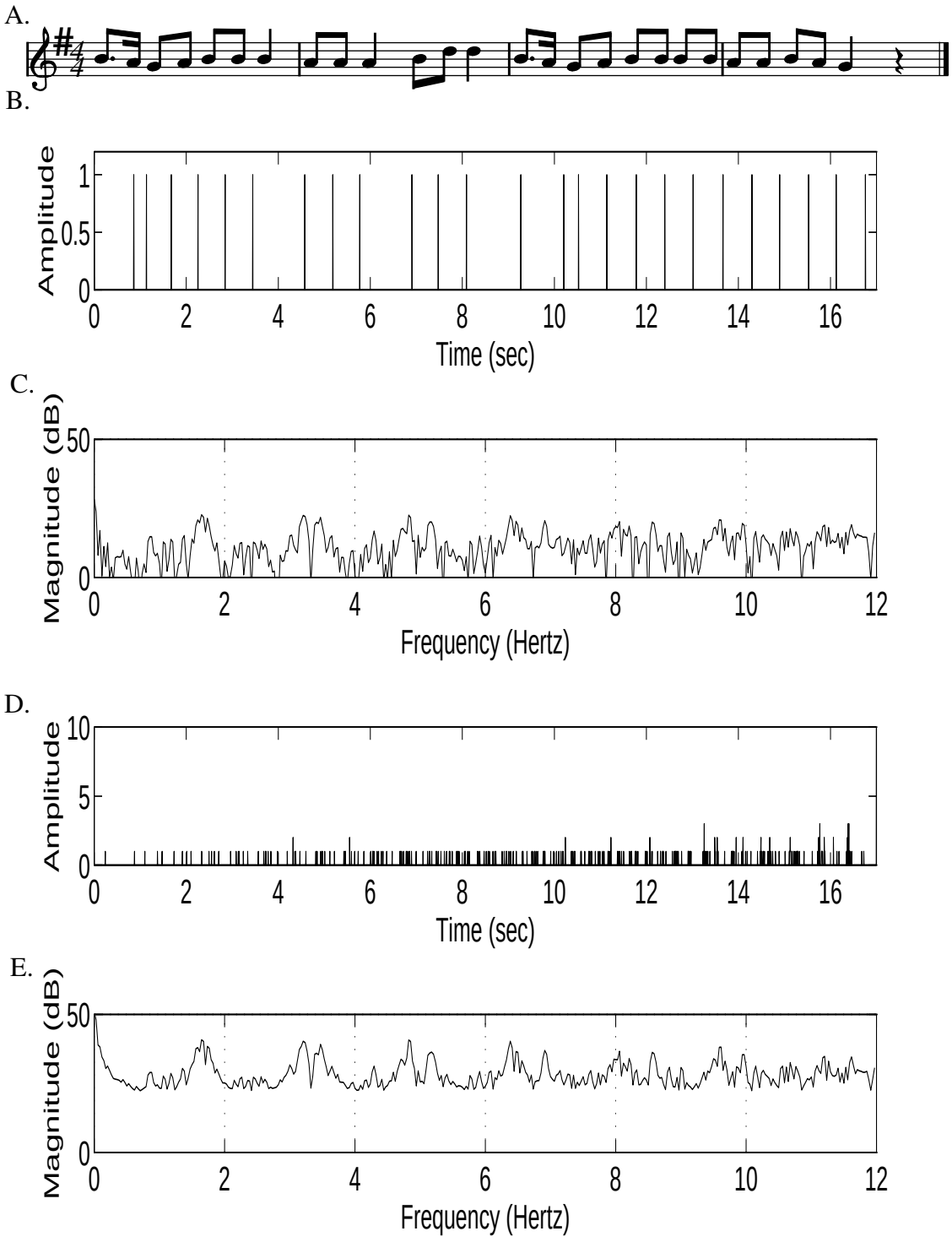


Figure 11: Performance of *Mary had a little lamb*: (A) musical notation, (B) onset times recorded in the performance, (C) discrete Fourier transform of B, (D) auto-correlation function of B, (E) discrete Fourier transform of D.

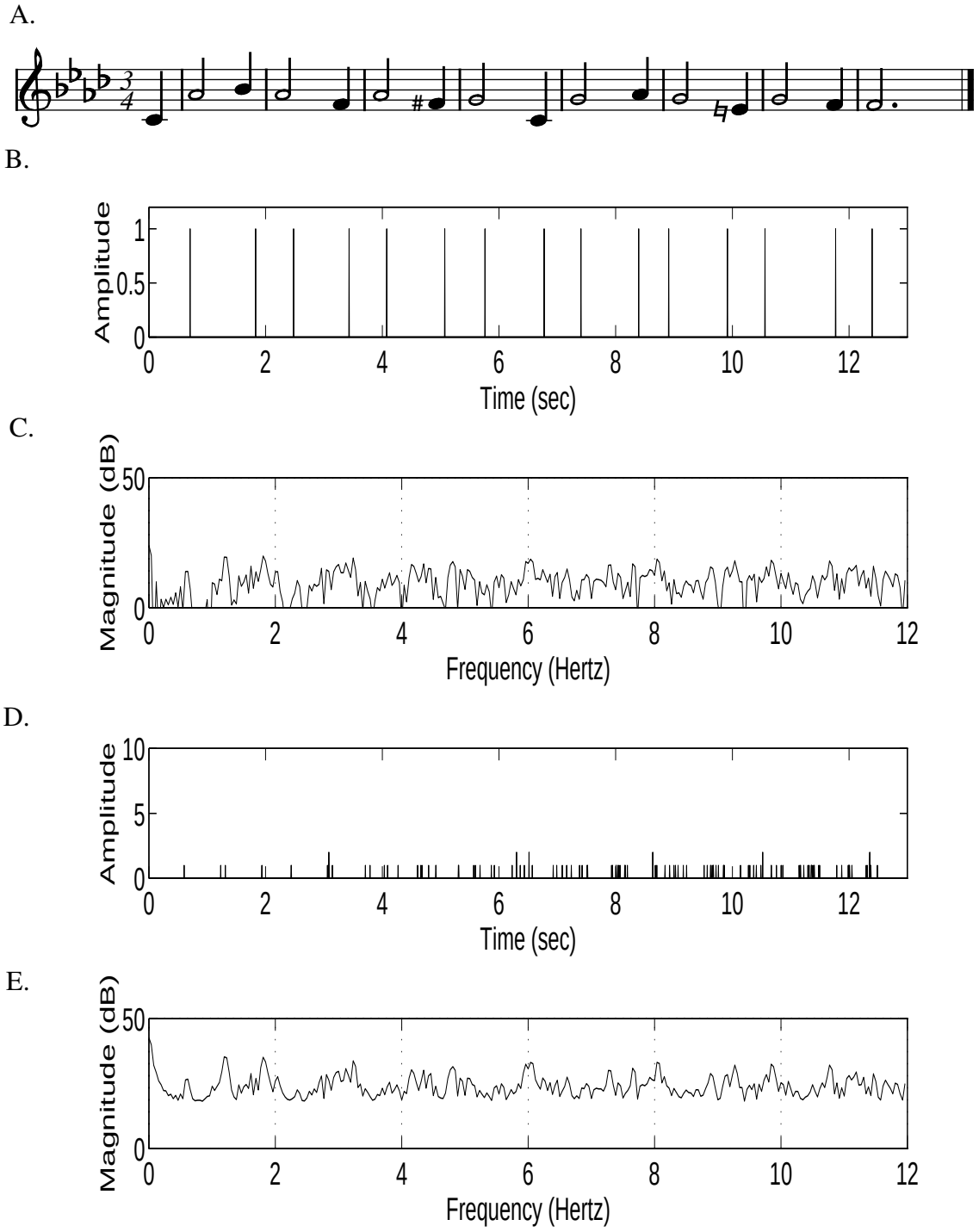


Figure 12: Improvisation on *Hush little baby*: (A) musical notation, (B) onset times recorded in the performance, (C) discrete Fourier transform of B, (D) auto-correlation function of B, (E) discrete Fourier transform of D.

3.5.2 Rubato

Rubato is deviation from temporal regularity, characterized by the shortening and lengthening of inter-onset intervals called for by the musical score (Palmer, 1988). For this study, as in previous studies of timing deviation (e.g. Bengtsson & Gabrielsson, 1983, Palmer, 1988), rubato was defined for each performed inter-onset duration as the deviation from the average duration, based on average tempo for each performance. To make comparisons across performances possible, deviations are expressed in percent.

To assess the amount and distribution of rubato among the performances a mean timing deviation was calculated for each performance. This measure of deviation was then averaged across the five performances of each melody or improvisation by each pianist. An analysis of variance (ANOVA) on mean deviation by performance type (melody/variation), subject, and tune was conducted. There was a significant main effect of performance type ($F(1,4) = 33.46$, $p < 0.01$), indicating that, on average, more rubato was used in the improvisation of variations than in the performance of the melodies from notation. Mean rubato was 0.05 for notated melodies, and 0.10 for improvisations.

There was also a significant interaction between melody and subject ($F(2, 8) = 13.89$, $p < 0.01$), as shown in Figure 13. Pianist 1 performed the melodies and improvisations for the first two tunes with little rubato, but for the third tune with high rubato. Pianist 2 performed tune three with little rubato, and performed tunes one and two with relatively high rubato. An addition, there was a significant three-way interaction between performance type, tune and performer ($F(2, 8) = 10.20$, $p < 0.01$) that is harder to summarize. Overall, however, the results suggest that performers exercised a great deal of control over the amount of deviation from ideal timing used in any given performance.

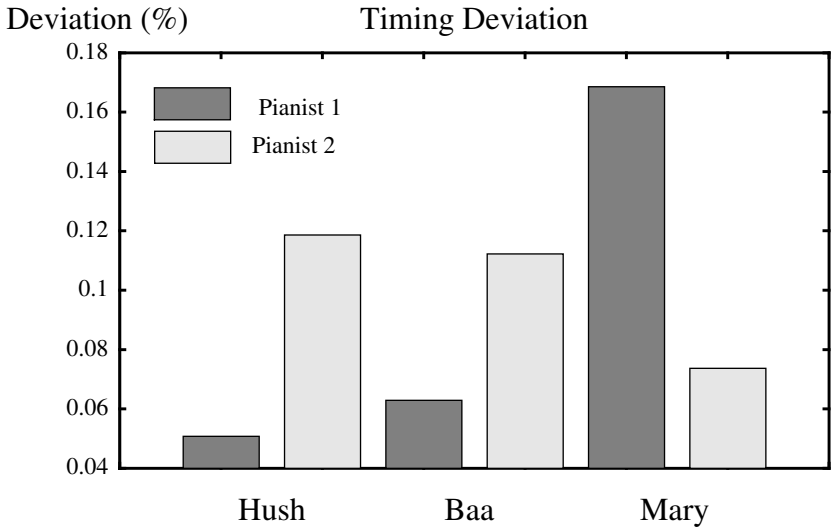


Figure 13: Bar plot showing a significant two-way interaction of subject and melody in deviation from average tempo.

3.6 Discussion

Analyses of timing in the performance of notated melodies and the improvisation of melodic variations revealed intricate temporal structure. Signals that reflected ideal patterns of relative timing produced Fourier spectra and auto-correlation functions that exhibited distinct patterns reflecting metrical structures. Fourier analysis and auto-correlation of *actual* performances and improvisations revealed temporal structures that were complex by comparison. The reason for this discrepancy is that the input signals corresponding to the performance data were nonstationary. Fourier analysis, for example, is most useful for identifying components of a signal whose phases and periods are fixed. One may even presume that a stochastic source (i.e. a random variable with some probability distribution) governs deviations from a given fixed generator, and Fourier analysis will produce good results: if the phase and period of the signal vary according to a fixed distribution there remains a central tendency for the spectral analysis to identify. However, performed metrical rhythms differ from stationary signals in that one cannot assume a fixed model of their generation. Neither can one assume a stochastic source, because usually timing deviations are systematic and carry information (e.g. Palmer, 1988).

Further analysis considered deviations from *ideal* patterns of timing for the performed melodies and the improvised variations. The most reliable predictor of variance was performance type, notated melody vs. improvisation. Although no attempt was made to identify the source of the timing deviations in the improvisations, listening to the improvisations reveals three possibilities. First, the pianists seem to have played the improvisations more expressively than the notated melodies. Second, in certain performances by both pianists, there are slight but audible pauses in which the performers seem to be deciding what to do play next. It is difficult to qualify this type of pause, except as a deviation from timing regularity. Third, in two improvisations there are actual mistakes: the pianist first played a wrong note and then either corrected himself, or paused very briefly before continuing. It is more difficult to know what to do with this type of timing variation. It is tempting to remove this data from analysis for several reasons. The most obvious of these is that the analysts are essentially forced to guess what the performer intended to play. Thus, part of the variance in these cases may be introduced by the analysis process. I chose to retain this data mainly because such mistakes are a valuable source of real-world performance noise in the data set. Although this complicates the interpretation of statistical analyses, it is precisely the type of data on which to test robustness of proposed models (see Chapter VIII). These discrepancies and analytical difficulties emphasize not only the challenge of explaining meter perception, it also emphasizes the difficulty of the task that humans are faced with in perceiving the meter of live musical performances.

CHAPTER IV

COMPUTING REDUCED MEMORY REPRESENTATIONS

4.1 Connectionism and Reductionist Music Theory

This chapter describes a model of sequence representation that is sensitive to structural relationships among events. More specifically, the model described here computes the relative importance of events in a musical sequence, consistent with a reductionist music theory (Lerdahl & Jackendoff, 1983). One difficulty with designing a mechanism specifically based upon Lerdahl and Jackendoff's (1983) theory lies in the specification of a relative weighting scheme for the set of rules that create reductions. A scheme has not yet been proposed that will work for every musical context. For complex musical pieces, one must enlist the aid of musical "common sense" in providing the proper weighting of musical considerations. A second problem regards learning. Reductionist theories assume that a great deal of musical knowledge is acquired as a result of experience with the musical culture or style in question. Empirical evidence suggests that a restructuring of mental representations for novel musical sequences may occur with as few as five or six exposures to a sequence (Serafine, Glassman, & Overbeeke, 1989). However, reductionist theories have not yet addressed the issue of how the musical knowledge necessary for the production of reduced descriptions is acquired.

An approach that offers a solution to these problems is the application of connectionist models, which learn internal representations in response to the statistical regularities of a training environment using general-purpose learning algorithms such as back-propagation (Rumelhart, Hinton, & Williams, 1986). The solution for musical variation offered by reductionist theories requires the representation of constituent structure, however, and connectionist models have been notoriously weak at representing constituent relationships such as those in language and music (Fodor & Pylyshyn, 1988). One approach to this problem involves learning distributed representations for compositional data structures using a recursive encoder network. This connectionist architecture, known as Recursive Auto-Associative Memory (RAAM), has been used to model the encoding of hierarchical structures found in linguistic syntax and logical expressions (Chalmers, 1990; Chrisman, 1991; Pollack, 1988; 1990).

To produce a memory representation for a musical sequence with the RAAM architecture, the sequence is first parsed to recover a compositional data structure that captures the sequence's time-span segmentation. A RAAM network can then be trained to produce a distributed representation for each time-span described by this structure. For example, the sequence of musical events in Figure 14, "a b c", may be represented as the nested structure ((a b) (c null)). A

compressor network is trained to combine a and b into a vector R1, to combine c and null into a vector R2,and then to combine the vectors R1 and R2 into a vector R3. A reconstructor network is trained to decode the vectors produced by the compressor into facsimiles (indicated by the symbol ') of the original sets of patterns. In the example, the reconstructor decodes R3 into R1' and R2', R1' into a' and b', and R2 into c' and null'. Thus, the vector R3 is a *representation* for ((a b) (c null)) because a reconstruction algorithm can be applied to R3 to retrieve a facsimile of the original sequence. It is a *distributed* representation because it is realized as a pattern of activation. It is a *recursive* distributed representation because its construction requires the network to recursively process representations that it has produced. The representations are *reduced descriptions* of musical sequences because the vector representation for an entire pattern is equal in size to the vector representation of a single event.

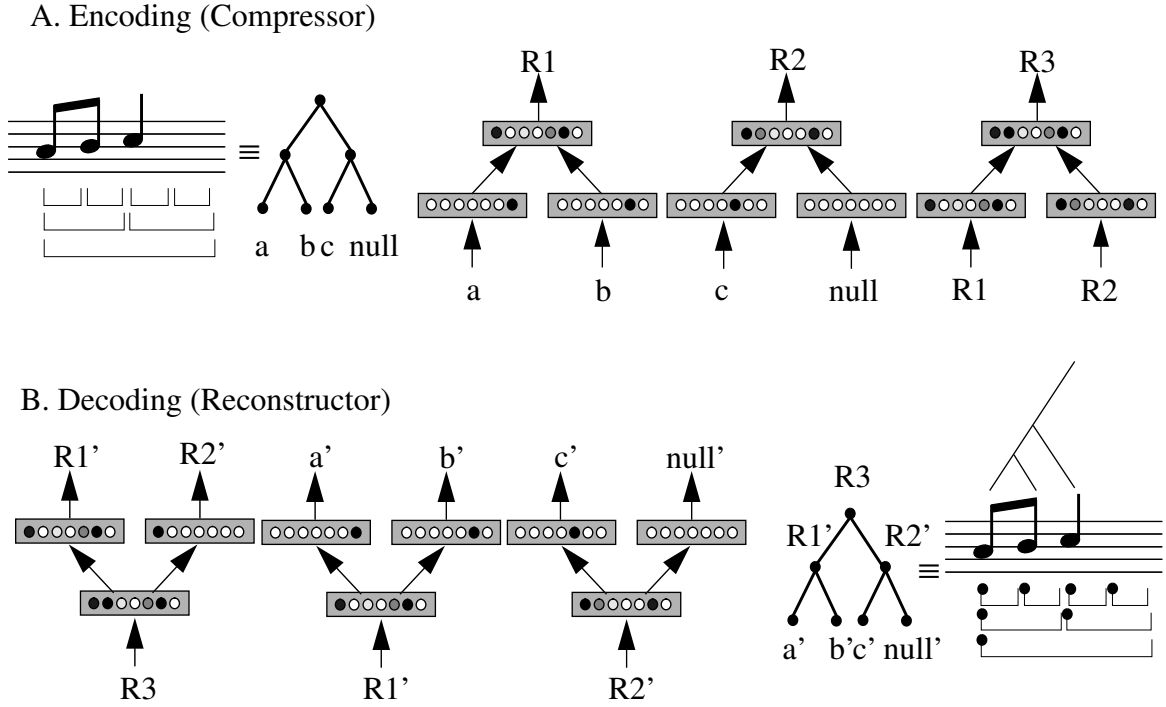


Figure 14: Encoding and decoding of a musical sequence by a RAAM network. (A) Based on a vector representation for each event and a constituent structure analysis, the compressor combines the group (a b) into a single vector, R1, (c null) into the vector R2,and then combines (R1 R2) into the vector R3. (B) The reconstructor decodes the vector R3 to produce (R1' R2'). It then decodes R1' to produce the facsimile (a' b') and R2' into (c' null').

The structures that the RAAM reconstructs are facsimiles of the original structures because the construction of a recursive distributed representation is a data compression process, which necessarily loses information. The network may reconstruct some events with lowered activation, and may fail to reconstruct other events entirely. The important question regards which events will be reconstructed faithfully, and which will be lost or altered in the compression/reconstruction

process. If, in the compression/reconstruction process, the network consistently loses information about less important events and retains information about more important events (i.e. as predicted by the music-theoretic analyses), then the network has also captured information that extends beyond pitch and time-span segmentation. Can the network training procedure discover the relative importance of events corresponding to metrical accent and time-span reduction?

If so, this supports the notion that reductions of musical sequences may be computed by a memory coding mechanism whose purpose is to produce descriptions for musical sequences that reduce demands on memory while retaining the gist of the sequences. This implies that the culture- and style-specific musical knowledge necessary for computing reductions is realized as a set of parameters (in a RAAM network, a set of weights) in the coding mechanism. The acquisition of this set of parameters can be viewed as the acquisition of the musical knowledge for computing reductions.

This view of reduced memory representations for musical sequences has several advantages over other possible mechanisms. The vector representations produced by a RAAM for melodic segments are reduced descriptions of the sequence, similar to the "chunks" proposed by recoding theorists. However, the compressed representation for a sequence is more than just a label or pointer to the contents of a structure (cf. Estes, 1973); it actually *is* the description of its contents. Therefore, the numeric vectors produced by the network potentially contain as much information as the cognitive structures proposed by pattern-formation theories. Because the reduced descriptions are represented as vectors, they are suitable for use with association, categorization, pattern-recognition and other neural-style processing mechanisms (Chrisman, 1991). Such processing mechanisms could, for example, be trained to perform sequence extrapolation tasks (Simon & Sumner, 1968).

This chapter describes two experiments with the Recursive Auto-Associative Memory (RAAM) architecture for producing reduced memory descriptions of musical sequences. RAAM networks are trained on a corpus of simple melodies and then tested in two ways. First, the networks' abilities to accurately compress and reconstruct a test set of three tunes are examined. In the second experiment, the structure of the representations produced by the network is also examined.

The network experiments have two goals. The first is to measure the performance of the RAAM networks using a *well-formedness* test (Pollack, 1990). For a given input melody, the compressor network creates a reduced description. The reconstructor network is then applied to the reduced description to retrieve its constituents. If the reconstructed sequence matches the input melody, either exactly or within some tolerance, then the reduced description is considered to be well-formed. The well-formedness test can also be used to measure the ability of RAAM networks to generalize, by testing the network's performance on novel sequences. The performance of the networks on a test set of three melodies is examined: *known*, *variant*, and *novel*. The *known* melody is one of the melodies presented to the networks during a training phase. Performance on this melody establishes a baseline of the networks' abilities to correctly encode melodies. The *variant* melody is a variation of material presented to the networks in the training phase, and the *novel* melody is a melodic sequence not related in any obvious way to the material presented in the

training phase. If the networks generalize from the examples presented in the training phase, then they should be able to produce well-formed reduced descriptions for one or both of the *variant* and *novel* melodies as well as for the *known* melody.

An additional goal in the second experiment is to determine the structure of the representations produced by the network. The network is provided with a time-span segmentation for each melody. The question is: Will the network take advantage of this information about temporal structure to preserve musical regularities that are systematically related to this structure? RAAM networks learn a data compression algorithm tailored to the statistical regularities of a training set. Thus, if the training set is adequately representative of the statistical characteristics of simple Western tonal melodies, the network should make use of this information, displaying significant levels of agreement among network, theoretical, and empirical measures.

4.2 A RAAM Architecture for Music

RAAM uses a connectionist substrate of fully-connected feed-forward neural networks to produce recursive distributed representations (Pollack, 1990). For example, to encode binary trees with k-bit patterns as the terminal nodes, the RAAM compressor would be a single-layer network with 2 k-unit input buffers and one k-unit output buffer. The RAAM reconstructor would then be a single-layer network with one k-unit input buffer and 2 k-unit output buffers. The input and output buffers are required to be the same size because the network is used recursively: the output of the network is fed back into the network as input. During training, the compressor and reconstructor are treated as one standard three-layer network ($2k$ inputs, k hidden units, and $2k$ outputs) and trained using an auto-associative form of back-propagation (Rumelhart, Hinton, & Williams, 1986; Cottrell et.al., 1988), in which the desired output values are simply the input values. To create the individual training patterns for the network, the structures that make up the training set are divided into groups (for example: (a b) or (R1 R2)).

Two issues arise in designing a RAAM network for encoding musical structure. First is determination of a constituent structure for each musical sequence that will specify how events are presented to the network input buffers. For these experiments, a time-span segmentation (Lerdahl & Jackendoff, 1983, see Section 2.2) of the melodies is used. For the simple melodies used in this study, the time-spans at smaller constituent levels (less than a measure) were "regular" (Lerdahl & Jackendoff, 1983); that is, they were aligned with the locations of strong metrical beats. Therefore, the lower levels of time-span segmentation were determined by the metrical structure. Grouping rules (Lerdahl & Jackendoff, 1983) were used to determine time-span segments at constituent levels larger than the single measure. Each encoding produced by the network is the representation of a time-span (and its events) at some level in the time-span segmentation. Once an encoding has been produced, temporal information is implicitly managed by the recursive structure of the decoding process. As decoding proceeds, the output codes represent smaller and smaller time-spans (at lower and lower levels), until, at the termination of the decoding process, a single pitch event is output and the temporal location of that event is uniquely determined.

Second, the representation of pitch events to be encoded by the RAAM must be specified. The different pitches in each melody are represented as binary feature vectors (on or off). A “local” representation of pitch class was used; 7 units represented the seven pitch classes of the diatonic scale in Western tonal music. Two units were also added to represent melodic contour. One unit means 'up' from the previous event, the other means 'down', and turning both units off means no contour change. This representation nominally captures octave equivalence and pitch height but makes no further assumptions regarding the psychophysical components of pitch, as other connectionist researchers have done (cf. Mozer, 1991). More sophisticated encoding strategies may prove useful for certain musical applications (cf. Large, Palmer, & Pollack, 1991), but this study sought to reduce inductive biases that would be introduced by more complex coding schemes.

Two modifications to the RAAM architecture are necessary to encode Western tonal melodies such as those in the training set. First, existing applications of the RAAM architecture have only accurately handled tree structures that are 4-5 levels deep. However, the 25 training melodies used in this study contain constituent structure hierarchies 6-7 levels deep, which expand to more than 1000 individual training patterns. Previous experiments found that this training set size outstrips the capacity of a RAAM network that contains a reasonably small number of hidden units (Large, Palmer, & Pollack, 1991). A method was adopted of scaling up the basic architecture by having one RAAM network recursively encode lower levels of structure, and then passing the encodings it produces to a second RAAM that encodes higher levels of structure. This method, known as *modular* RAAM (Angeline & Pollack, 1990; Sperdutti, 1993), enables the construction of recursive encoders that can handle trees with many hierarchical levels by using multiple networks that each contain fewer hidden units. To use this strategy, a smallest possible time span was identified, corresponding to the smallest duration value that occurred in the network training set. Thus the training set consisted of balanced trees. The lower RAAM network was trained on a fixed number of nested levels, and the upper RAAM was trained on the upper levels of the trees.

The second modification addresses ternary groups common in Western tonal music; binary branching structures are not sufficient to capture musical groupings that often consist of three elements. To handle both pairs and triples, a network with three input buffers might be created, only two of which would be used to encode binary segments. However, this would lead to the situation shown in Figure 15A, in which a triple with a rest in the middle is indistinguishable from a pair. Instead, a network with four input buffers can encode both duple and triple segments and distinguish among them, as shown in Figure 15B. Here buffer 1 corresponds to the first event of any group, buffer 3 corresponds to the second event of a binary group, and buffers 2 and 4 correspond to the second and third events, respectively, of a ternary group. To properly interpret the output of these buffers at decoding, four extra units were added at the output. The network is trained to turn on an output unit when the corresponding buffer's output is to be used; otherwise the contents of the buffer are ignored, and trained with a don't-care condition (Jordan, 1986).

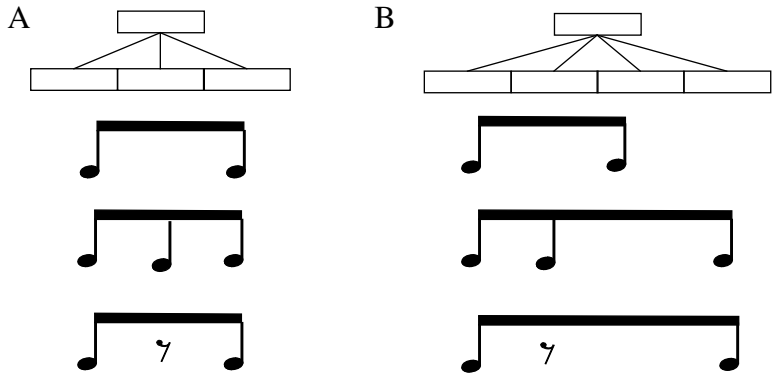


Figure 15: Buffering scheme for encoding either duple and triple grouping structures. (A) Three buffers cannot discriminate between a group of two events and a group of three events in which the middle event is a rest. (B) Four input buffers can make the discrimination.

4.3 Experiment 1: Balanced Tree Structures

4.3.1 Training

Twenty-five simple children's melodies were chosen as a training set because they provided a simple, natural musical case for study. The tunes comprised eighteen unique melodies; five of these eighteen melodies had variations in the training set. Each melody was between 4 and 12 measures in length, with a time-signature of 2/4, 3/4, 4/4, 6/8, or 12/8. The tunes provided constituent structures six to seven levels deep, in which either binary or ternary groups appeared at each level. Although the pitch event representations required only 9 bits (7 pitch class units and 2 contour units), 35 units were used, allowing 26 extra "degrees of freedom" for the system to use in arranging its intermediate representations. These extra dimensions of representation were set to 0.5 on input, and trained as don't-cares (Jordan, 1986) on output. As described above, the two RAAM modules each required four input buffers, and each resulting module had 140 input units, 35 hidden units, and 148 output units.

The first module was trained on the bottom 3 levels of the trees, such that the input corresponded to metrical levels up to and including coding of the "tactus" level. The representations that emerged from the lower RAAM (the output) corresponded to time-spans with a length of one half-note for binary groups, or one dotted half-note for ternary groups. The second module was trained on the upper 3 or 4 levels of the trees (depending on tune length), corresponding to larger structural levels of the melodies. This division of labor allowed the modular architecture to approximately balance the learning load between the two modules, measured by the number of unique training patterns. The two modules were trained simultaneously, with the bottom module's output providing the input for the top module. Rather

than exposing the network to the entire training set of twenty-five melodies, four melodies were chosen randomly from the training set and forward-propagated in each training cycle; then error was back-propagated through the network (cf. Cottrell & Tsung, 1991). This method allowed a faster running time for the large training set. Because the length of the tunes in the training set (and therefore the number of individual training patterns) varied for each cycle of backpropagation, the learning rate was set to 0.7 divided by the number of training patterns seen on that cycle. Momentum was set to 0.5 and weight decay to 0.0001. Training lasted for 1300 cycles, by which point the error associated with the test set of melodies reached a minimum value.

4.3.2 Testing

In this experiment, performance was judged based on the well-formedness test, which assessed the ability of the network to accurately compress and reconstruct each melody. The network was tested on a set of three melodies: a *known* melody, a *variant* melody, and a *novel* melody. These were the same three melodies used in the empirical study of improvisation; the names used here denote the particular relationship of each melody to the network training set. The *known* melody, *Mary had a little lamb*, occurred in the training set. The network's performance on this melody is representative of the network's performance on familiar (learned) melodies. The *variant* melody, *Baa baa black sheep*, did not occur in the training set; however, four closely related variations of this melody did occur in the training set. The local structure (duration patterns and melodic contour of individual measures) of the variant melody was very similar to two training set melodies, and the global structure (three 2-measure phrases with similar melodic and harmonic implications) of the variant melody was similar to two other training set melodies. Performance on this melody indicates the ability of the network to account for simple melodic variation, because the network was required to recombine familiar local structures (individual measures) in novel global contexts (different melodies) that shared structural features with known melodies. The *novel* melody, *Hush little baby*, did not occur in the network training set, nor was it closely related to any of the melodies in the training set. Network performance on this melody indicates the ability of the network to perform a type of generalization different from that required for melodic variation: the ability to represent novel musical sequences at local levels of structure, as well as the ability to combine novel local structures in novel global contexts.

Each melody was reconstructed by the decoder network from the recursive distributed representation produced by the compressor. Errors in the reconstructed melody took the form of additions (the network reconstructed an event that was not present in the original melody), deletions (the network failed to reconstruct an event that was present in the original melody), and substitutions (the network reconstructed an event incorrectly in the same position). An error measure was created based on the number of sixteenth-note locations in each piece, because this was the smallest time-span in each segmentation. There were 64 (16x4) sixteenth-note locations in the known melody, 96 (16x6) in the variant melody, and 32 (8x4) in the novel melody. Given the coding scheme, the chance estimate for percentage of events correct at each location is 1/16, or 6.25%, based on 16 possible outcomes: seven pitch classes times two contour changes (up or down) plus a repeated pitch and a rest.

As an approximate measure of the network's ability to correctly compress and reconstruct constituent structures, average performance on the training set melodies was calculated. Performance was measured at two points in the time-span segmentation for each melody. First, the network's ability to compress and reconstruct time-span segments with only three levels of recursive nesting – corresponding to a time-span of one half note for binary groups – was examined. Network performance in reconstructing training set melodies with three levels was 84%. Next, the network's ability to compress and reconstruct time-span segments that corresponded to entire melodies, with 6-7 levels of recursive nesting, was examined. Here the network's performance was 57%. Thus, the representations captured lower-level structures fairly well, whereas at global levels of structure, the representations lost sequence details.

To better understand the network's performance, the reconstructions for the three test melodies were examined in detail, shown in Figures 16, 17, and 18. The reconstruction of the *known* melody shows network performance on melodies learned in the training set. Figure 16 shows reconstructions made from codes at each level of hierarchical nesting. At the lower levels the reconstruction is nearly perfect, but at five to six levels of nesting, the network has lost quite a bit of information.

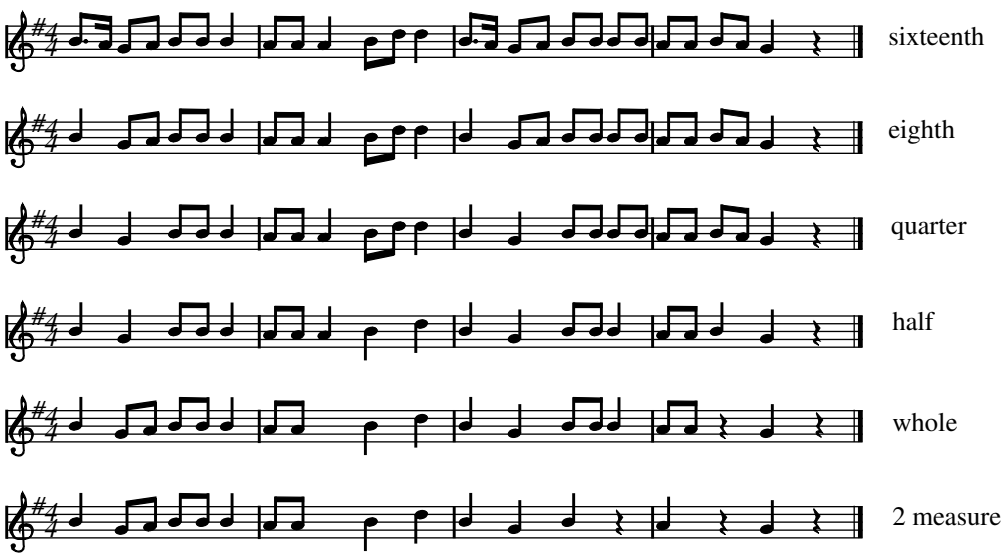


Figure 16: Network reconstructions of *Mary had a little lamb* in Experiment 1.

The reconstructions of the *variant* melody, shown in Figure 17, gives an indication of the network's performance on simple variations of learned melodies. Again, the reconstructions produced at lower levels of nesting are accurate, while considerable information loss occurs after several levels of recoding.

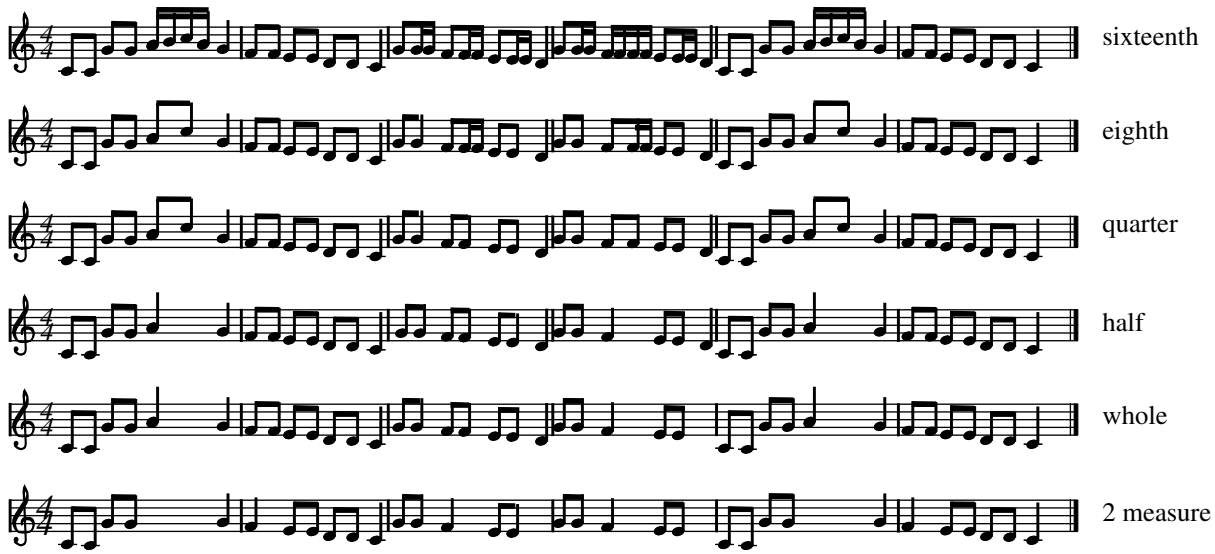


Figure 17: Network reconstructions of *Baa baa black sheep* in Experiment 1.

Finally, the reconstructions of the *novel* melody, *Hush little baby*, shown in Figure 18, give an indication of the network's performance on unlearned material. At lower levels, the network shows some ability to generalize, but at higher levels network performance is poor.

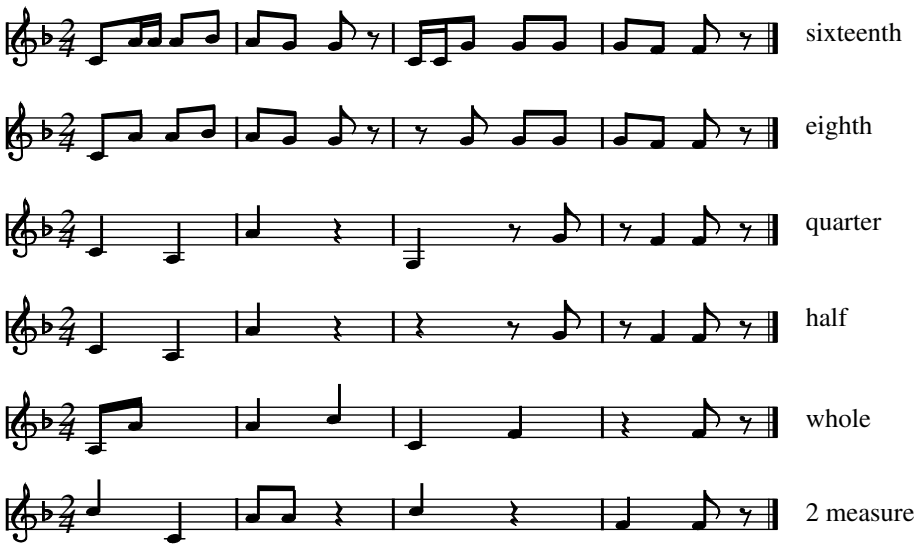


Figure 18: Network reconstructions of *Hush little baby* in Experiment 1.

4.4 Experiment 2: Unbalanced Tree Structures

The problem with the network of Experiment 1 was that it did not do a good job of representing the sequences at higher levels of recursive encoding. However, examination of the reconstructions reveals a result that *is* encouraging – the events reconstructed correctly by the network appear to be the more important events of these melodies. Thus there may be information in the training set that the network can extract, allowing it to identify structurally important elements. However, the network training strategy may be preventing it from accurately reconstructing the sequences. One difficulty may be the use of empty time-spans at lower levels of nesting forcing the network to spend resources to correctly reconstruct null events.

For the experiment, the training strategy and network architecture were adapted to code unbalanced trees. Rather than creating fully balanced trees for each melody, trees were constructed whose nesting levels corresponded only to those necessary for describing the sequence. This strategy is more closely analogous to Lerdahl & Jackendoff's (1983) time-span segmentation theory. This form of training, however, requires that it be clear when to terminate the decoding process (Pollack, 1990). This problem was addressed by adding an extra unit to each RAAM module, trained as a terminal detector. The terminal detector allows the network to determine a) when to pass a code from the higher-level RAAM network module to the lower-level RAAM module during decoding, and b) when to interpret a code produced by the bottom module as a pitch event. Knowing when to terminate decoding is equivalent to determining the level of the time-span segmentation to which a melodic event corresponds. This allows a more flexible strategy for coding time-span segmentation trees, shown in Figure 19, allowing the network to concentrate its resources where they are needed. However, it may also complicate the learning process, since in this strategy, for example, a quarter-note "C" is no longer literally similar to two eight-note "C"s in succession. The network will have to learn this for itself.

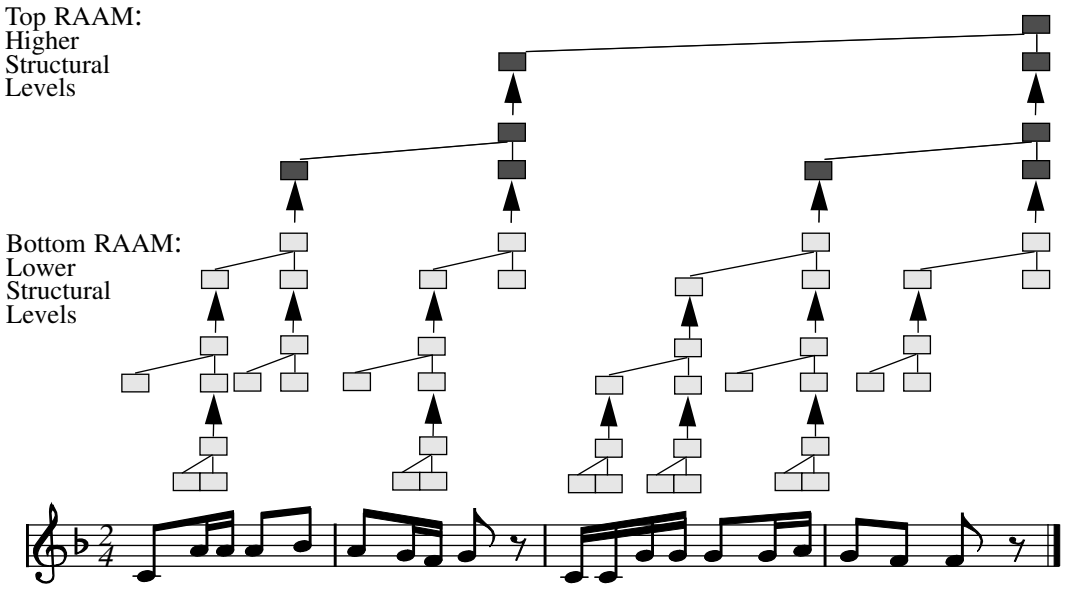
4.4.1 Training

The training materials and procedure from Experiment 1 were duplicated in Experiment 2.

4.4.2 Testing

In this experiment, the network's performance was measured in two ways. First, well-formedness tests assessed the ability of the network to accurately compress and reconstruct each melody, and revealed the basic representational capacity of the network. Second, tests of representational structure assessed the relative weighting of constituents on an event-by-event basis, and revealed the nature of the representational strategy developed by the network.

A. Encoding (Compression)



B. Decoding (Reconstruction)

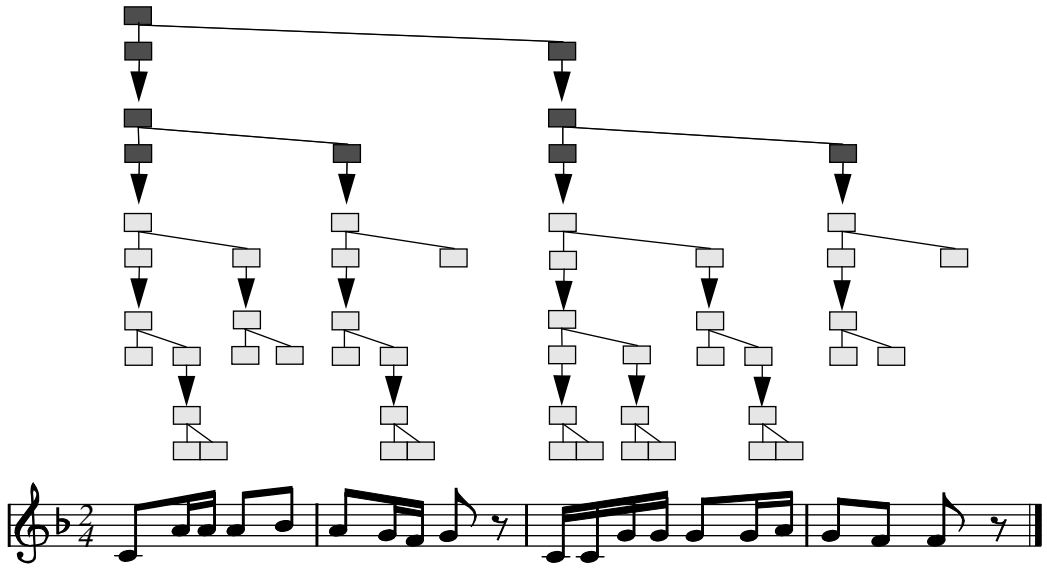


Figure 19: Simplified schematic of modular RAAM encoding used in experiment 2. (A) In the encoding diagram, time flows from left to right and bottom to top. (B) In the decoding diagram, time flows from top to bottom and left to right. The network determines when to stop decoding automatically.

4.4.2.1 Tests of Well-Formedness.

As a comparative measure of the network's ability to correctly compress and reconstruct constituent structures, average performance on the training set melodies was again calculated. Performance at two points in the time-span segmentation was measured for each melody. First, the network's ability to compress and reconstruct time-span segments with only three levels of recursive nesting was examined. Network performance in reconstructing training set melodies with three levels was 92%. Next, the network's ability to compress and reconstruct time-span segments that corresponded to entire melodies, with 6-7 levels of recursive nesting, was examined. Here the network's performance was 71%. The representations captured lower-level structures more faithfully, whereas at global levels of structure, the representation again began to lose sequence details, although loss was not as severe as in Experiment 1.

To better understand the network's performance, the reconstructions for the three test melodies were again examined in detail, which are shown in Figure 20. The reconstruction of the *known* melody reflects network performance on melodies learned in the training set. The reduced descriptions produced by the lower-level RAAM module were first examined (subsequences of events up to the level of half notes; lowest 3 levels of hierarchical nesting). In this reconstruction, the network made a single error, adding an event in the third measure, for a performance of 98%. Reconstruction at the whole tune level (all 7 levels of hierarchical nesting) resulted in four errors, giving an overall performance of 94% (60/64) for this melody, which was significantly better than chance (binomial test, $p < .01$). This reconstruction was better than the average for training set melodies, probably because two instances of this melody occurred in the training set.

The reconstruction of the *variant* melody gives an indication of the network's performance on simple variations of learned melodies. The reconstruction produced by the lower-level RAAM module for subsequences corresponding to half-notes (3 lowest levels of hierarchical nesting) resulted in performance of 92% (88/96). The network successfully learned the lower-level details because most of these surface features were present in the training set. The network's reconstruction at the whole tune level (all 7 levels of nesting) resulted in fifteen errors, for a performance of 84% (81/96), again significantly better than chance ($p < .01$). The reconstruction of this melody at (only) the whole-tune level was the same as its reconstruction of *Twinkle twinkle little star*, one of the four related melodies in the training set, for which its performance was 98% (94/96 events). As Figure 20 shows, the half-note level representations preserved local structure. The ability to exploit constituent structure, combined with the use of a recursive encoding strategy, allowed the network to rely upon structural similarities at the whole-tune level, rather than melodic and rhythmic features at lower levels, in determining the representation of this melody.

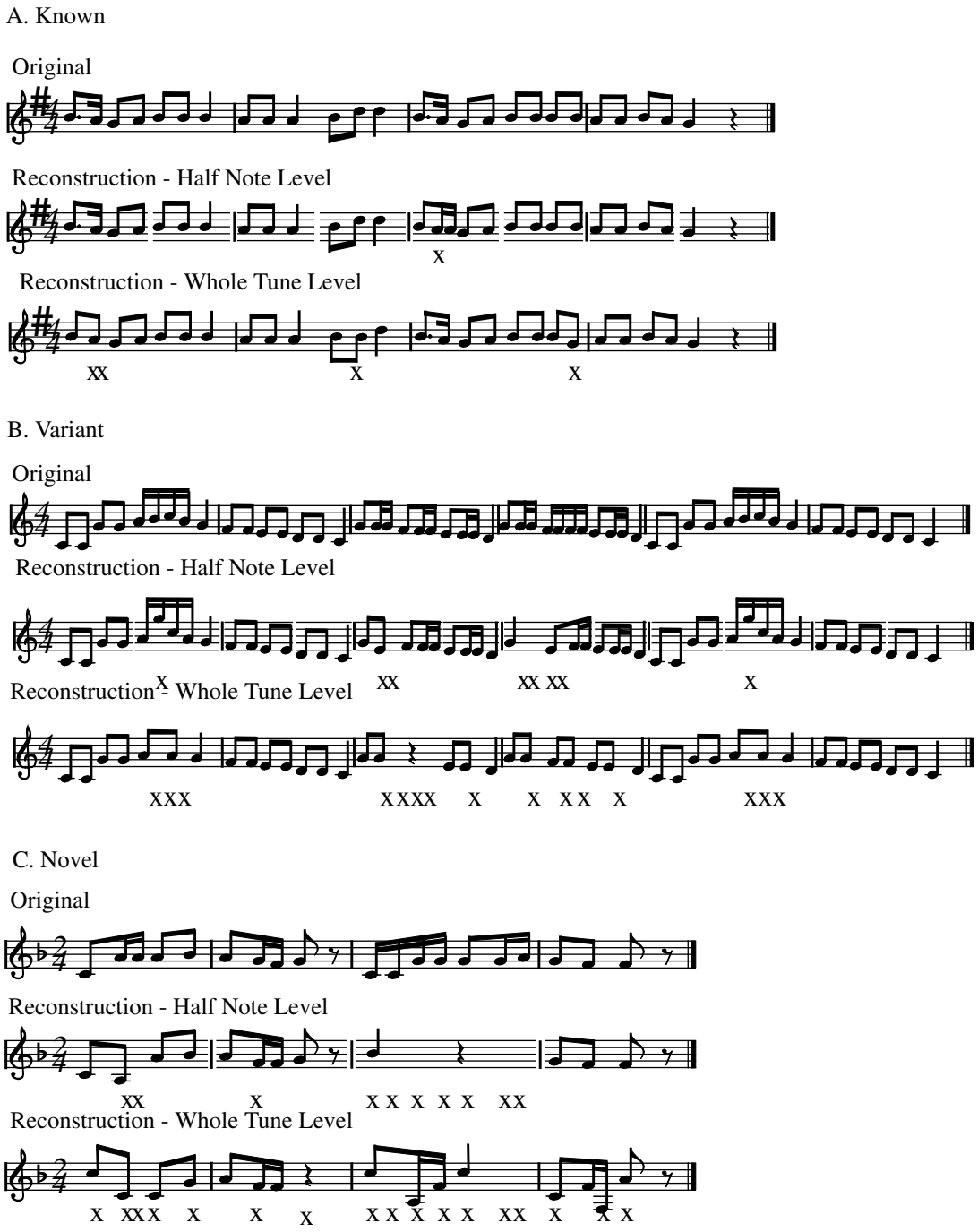


Figure 20: Original melodies and network reconstructions: (A) *Mary had a little lamb* (Known), (B) *Baa baa black sheep* (Variant), (C) *Hush little baby* (Novel). Each melody was reconstructed from several codes (the half-note level RAAM), and from a single code (the whole-tune level RAAM). X's denotes failures in network reconstructions.

The reconstruction of the *novel* melody is representative of the RAAM's ability to encode novel sequences. Again, the reduced descriptions produced by the lower-level RAAM module were first examined for subsequences of the original melody by encoding groups of events only up to the level of half-notes (3 levels of hierarchical nesting). Figure 20 shows the reconstruction from the reduced descriptions for each half-note of the tune. The lower-level reconstructions produced ten errors, for a success rate of 69% (22/32), again significantly better than chance ($p < .01$). Seven of the ten errors occurred in the third measure, and the other three measures of the tune were reconstructed rather faithfully. At the whole tune level there were seventeen errors, for a performance of 47% (15/32), which is significantly better than chance ($p < .01$), but overall, the reconstruction is poor (there are only 19 events in the original tune). It is interesting that the rhythm was reconstructed well (27/32, or 84%), but very few pitch events were reconstructed correctly (3/19, or 16%). Thus the network's representation of this melody at the whole tune level was not well-formed, and generalization to this novel sequence was better at the lower levels of the hierarchy.

4.4.2.2 Tests of Representational Structure

Given the network's relative success at representing the sequences, the structure of the distributed representations was analyzed to determine the relative contributions of individual events. One method is to directly examine the representation vectors to determine the function of individual hidden units. Little information can be retrieved from recursive distributed representations of this size, however, because of their complexity (Pollack, 1990).

As an alternative approach, the "certainty" with which the network reconstructed each event of the original sequence was measured by computing the distance between the desired (d) and obtained (o) vector representations at each sequence location (i). This analysis considered only those output units that represent pitch class (ignoring contour), consistent with the analysis of the improvisations. To compensate for the fact that some events were added and others deleted in the reconstructions, only the locations in the reconstructions for which pitch vectors should have been output were considered. Thus, only deletions and substitutions of events from the original melody affected this measure, as in the empirical study (Chapter III). A similarity measure was defined that ranged from 0 (most different) to 1 (identical), and represented the probability that desired pitch events occupied the appropriate positions in the original sequence, based on the network representation.

$$sim(d, o) = 1 - \left(\sqrt{\sum_{i=1}^{n} (d_i - o_i)^2} \right) / n$$

Sequence locations at which this measure was smallest were locations at which the network was most likely to make a reconstruction error. These probabilities were then interpreted as predictions of relative importance for each event in the distributed representation.

The probability measures of relative importance at the whole-tune level were correlated with the musical improvisation data, as summarized in Table 2. The correlations were large for the

known ($p < 0.10$) and variant ($p < 0.05$) melodies, but not for the novel melody. This was not surprising because the novel melody did not have a well-formed distributed representation at the whole-tune level. However, when the novel melody was reconstructed from the reduced descriptions corresponding to the half-note level of the tune (shown on the bottom of Figure 20), the resulting correlation approached significance ($p < 0.10$).

Table 2: Squared correlation coefficients for network reconstructions.

	Known (*Mary*) Whole Tune	Variant (*Baa*) Whole Tune	Novel (*Hush*) Whole Tune	Novel (*Hush*) Half-Note
Improvisation Data (# events retained)	.35*	.64**	.10	.40*
Metrical Accent Predictions	.39**	.55**	.24	.45**
Time-Span Predictions	.39**	.64**	.25	.52**
Semi-Partial (metrical accent removed)	.14	.35**	.27	.29

*-$p < 0.10$

** - $p < 0.05$

Next, the network measures of relative importance were compared with the quantifications of theoretical predictions, as shown in Table 2. The correlations with time-span reduction predictions were significant for the known and variant melodies and for the measure-level reconstruction of the novel melody ($p < 0.05$) but not for the whole-tune-level reconstruction of the novel melody. The correlations with metrical accent predictions also were significant for each melody ($p < 0.05$). The network measure was correlated with time-span reduction predictions after metrical accent was partialled out. The semi-partial correlation was not significant for the known or novel melodies, but was significant for the variant melody ($p < 0.05$), indicating some ability of the network to extract structure beyond metrical accent.

4.5 Discussion

Recursive Auto-Associative Memory produced recursive distributed representations for musical sequences. A general learning algorithm, backpropagation, extracted sufficient information from a

training set of 25 simple melodies to produce reduced descriptions of known, variant, and novel sequences. The performance of both networks was investigated using the RAAM well-formedness test. In each case, the network failed to reconstruct some events, reconstructed other events incorrectly, and occasionally added some that were not present in the original sequence. The second network performed better, because it coded melodies as unbalanced trees.

The reconstructions of melodies produced by the second network were fairly accurate, but did not retain all of the details. The network produced reduced memory representations that preserved the important structural features of the sequences. However, three sources of evidence suggested that the representations successfully captured the major structural features of the melodies. First, the reconstructions were faithful to the rhythm of the original melodies, even for the novel melody. Second, the network correctly reconstructed most of the pitches in the original melodies. Third, the events on which the network made reconstruction errors tended to be the less important events, as shown by the correspondence of network predictions of relative importance with theoretical predictions and improvisational data.

The network performed best on familiar (learned) melodies. The ability of the network to generalize was also tested: to represent both a variant of a learned melody and a truly novel melody (one unrelated to the learned melodies). The performance of the network in reconstructing the variant melody showed how the network handles simple melodic variation. This melody shared local structure with training set melodies, and the network's lower-level codes (up to the half-note level) preserved this structure. At a global level, the compression/reconstruction process followed the attractor (a known path) for another melody with which the variant shared global structure. The network also identified the important pitch events in the variant, indicated by the fact that network measures of relative importance for this melody correlated strongly with the time-span reduction predictions. Comparison with the empirical data from improvisations supported the conclusion that the network successfully identified events interpreted as major structural features by musicians. Overall, these results demonstrate the ability of the network to exhibit a limited but important form of generalization.

The findings for the novel melody indicated that the network still performed well at lower levels of structure in handling unlearned sequences; it produced well-formed memory representations for the three lowest levels of the constituent structure. At higher levels of structure, however, the network failed to generalize, reproducing the correct rhythm but incorrect pitches for this melody. This aspect of performance may be due to the learning environment, which may not have provided a rich enough set of patterns at higher levels of structure.

The information retained by the network in the compression/reconstruction process agreed well with music-theoretic predictions of the relative importance of musical events. The limited size of the training and test sets make it difficult to say precisely why the agreement occurred; however, the time-span segmentation used as input to the network was related to the music-theoretic predictions. The network used this information about rhythmic structure, coded as position in a fixed input buffer, to learn representations that retained musically important events and major structural features. For instance, the network may have learned metrical accent by weighting the first element of lower-level time-spans (which aligned with strong beats) more heavily than others.

The relative importance predictions, however, were based on more complex rhythmic relationships. To learn relative importance, the network learned other stylistic factors. For example, the RAAM network learned that the last event in each sequence was predictable - it was always the tonic. Thus, the network appears to have extracted some relationships beyond metrical accent, and did so strictly on the basis of the regularities in the training set. The network was forced to distill musical regularities such as these from the training set in response to two opposing pressures: 1) to retain as much information about each sequence as possible, and 2) to compress the information about each sequence into a pattern of activation over a small number of units.

Finally, the psychological plausibility of this approach to creating reduced memory representations for music was demonstrated. Certain events dominated the structure of the reduced descriptions by virtue of the fact that they had the greatest probability of being correctly reconstructed by the network. The events that dominated the network's reduced descriptions were precisely those events most important in the mental representations for these melodies measured by the musical improvisations and posited in the theoretical reductionist predictions. These findings indicate that the RAAM coding mechanism produced reduced descriptions for musical sequences that implicitly weighted events in each sequence in terms of their relative structural importance. This is an important finding because it supports the psychological plausibility of recursive distributed representations as an approach to modelling human memory. Combined with the network's performance in reconstruction, these findings suggest that the reduced memory representations successfully captured the structure of musical sequences in ways similar to the mental representations underlying improvisational music performance.

CHAPTER V

SEQUENCE PROCESSING AND TEMPORAL PROCESSING

The RAAM network of the previous chapter did a good job of capturing the sequential structure of musical melodies. It represented sequences with long distance dependencies. It also generalized well enough to capture relative importance among musical events. The network did this because of the way it exploited information about relative timing that was available as input. However, this raises a difficult question: If timing in music is as flexible as shown in Chapter III, how can such relative timing information be made available to a sequence processing network? This chapter attempts to answer this question. First, it proposes a distinction between *sequence processing* and *temporal processing*. Next, the RAAM implementation is compared with other temporal sequence processing networks with respect to the handling of sequential relationships and temporal relationships. It is proposed that the processing of temporal structure is a key factor in the performance of real-time temporal sequence processing architectures. Previous models of temporal structure processing are reviewed and an entrainment model is proposed for handling temporal relationships in the processing of temporal sequences.

5.1 Temporal Sequence Processing

A temporal sequence can be notated as: X = [A^{200} B^{200} C^{400}]. According to this notation, each sequence element consists of a letter representing a sequence element, and a superscript representing event duration. Within a particular domain such as music, specific commitments must be made, for example letters may represent pitch events and the superscripts may represent inter-onset-durations. However, this is a general notation that may be used to describe any temporal sequence. The sequence X may be decomposed into two plain sequences, S = [A B C] and T = [200 200 400]. S is a sequence of elements; time is abstracted away and what remains (of time) is serial order. S is a pattern *in* time. T = [200 200 400] is a plain sequence of intervals; the events have been abstracted away and what remains (of the elements) is the way they have structured the temporal continuum. T is a pattern *of* time. The questions of temporal sequence processing may be likewise decomposed. One set of questions regards how a system handles sequence structure. Another set of questions regards how the system handles temporal structure. A third set of question regards how the system handles the interaction between sequence structure and temporal structure.

Within the artificial neural network community, a great deal of attention has focused upon questions of sequence structure. Two important issues have been studied: the design of short term memory (Mozer, 1993; Wang, in press a), and ability to make generalizations about sequence

structure (de Vries & Principe, 1992; Giles, et. al. 1990; Kolen, 1994; Pollack, 1988; Pollack, 1991). The goal of short term memory (STM) is to retain relevant aspects of the sequence history so that processing (sequence prediction or recognition, for example) may proceed. An important question regards the adequacy of STM design for this task, because relevant relationships among sequence elements often span long temporal intervals and involve high-order statistics (Elman, 1990; Jordan, 1986; Wang & Arbib, 1993; for reviews see Mozer, 1993; Wang, in press a). Next, sequence processing systems may need to make generalizations about sequence structure, for example to learn the structure of naturally or artificially generated languages (e.g. de Vries & Principe, 1992; Giles, et. al. 1990; Kolen, 1994; Pollack, 1991; Cleeremens, Servan-Schreiber, & McClelland, 1989). Generalization is important in learning musical styles and recognizing musical variation. Important questions regard the nature of the task specification and the nature of the processing algorithm (Wang, in press a).

Less work has focused directly upon questions of temporal structure. An important issue regards the design of rate-invariant systems. Systems for processing music and speech, for example, must process sequences independent of absolute presentation rate, yet maintain sensitivity to certain relative time relationships. These issues are related to a problem known as the *quantization*, or *time-warping* problem. This problem is difficult because there is a trade-off between relative-time sensitivity and rate-invariance: To what relative-time relationships should processing be sensitive, and to what other aspects of timing should processing be invariant? Other questions regard how systems make use of relative timing relationships and structures. This section compares RAAM with other temporal sequence processing architectures, observing this distinction between sequence processing and temporal processing.

5.1.1 Sequence Processing

This section explores a variety of temporal sequence processing architectures, and discusses how each addresses the problems of sequence structure as defined above. Two important issues are addressed. The first issue is the design of short term memory (STM). Form, content, and adaptability of memory structures (Mozer, 1993) are discussed. The discussion of STM structure suggests that many temporal sequence processing architectures address the role of time in sequence processing in a limited way: time enters the picture as a constraint on the maintenance of sequence history in STM. The ability of each architecture to make domain-relevant generalizations is also evaluated. Thus, this review concentrates on musical applications where possible, as a way of evaluating generalization potential. The discussion of generalization shows that temporal sequence processing systems that incorporate the temporal structure of sequences directly into processing generalize best in the musical domain.

One temporal sequence processing architecture studied extensively is the time delay neural network, or TDNN (e.g. Elman & Zipser, 1988; Waibel, et. al. 1989; for comprehensive reviews see Mozer, 1993; Wang, in press a). The TDNN takes its name from the structure of its short-term memory store: STM makes a subset of past events simultaneously available for processing using a set of tapped-delay lines. Processing is most often accomplished using a multi-layer perception trained with backpropagation. In the design of STM care must be taken that there are enough lines

with proper delays to provide adequate context for the current task. In the simplest strategy, delay line structure is fixed, imposing a strict upper limit on the number of items that can be held in STM at once, and the N most recent inputs make up the contents of the memory.

The RAAM network discussed in the previous chapter is a type of delay line network. However, the RAAM strategy for making use of delays was complex. The content of memory consisted not only of past sequence elements but also of recoded chunks (see Figures 14 and 19), capturing the sequence history in a more powerful way. Each delay buffer could hold either a single past sequence element or a chunk that captured a larger amount of sequence context. This strategy also required the use of an external stack to handle intermediate results. In addition the RAAM implementation made sophisticated assumptions regarding the handling of time as will be discussed in the next section. As already discussed, the RAAM generalized well in the musical case, capturing psychologically relevant structural relationships among sequence elements.

Another popular STM design strategy is the exponential trace memory, studied by Jordan (Jordan, 1986), Mozer (Mozer, 1989) and Wang and Arbib (Wang & Arbib, 1990), among others. In its simplest form, an exponential trace STM consists of a decaying trace of past sequence elements. This STM design does not impose a fixed limit on the number of past events that can affect the current processing as in the time-delay strategy. In practice, however, the number of past events that affect processing is usually small.

The exponential trace STM has been widely studied in music processing in conjunction with a variety of different tasks and processing strategies. Many researchers, for example, have explored musical structure using discrete-time recurrent networks trained with backpropagation (e.g. Bharucha & P. Todd, 1991; Burr & Miyata, 1993; Mozer, 1991; Mozer, 1994; Stevents & Wiles, 1994; P. Todd, 1991). The task of a recurrent network (RN) is usually to predict events in a sequence, thus RN's provide natural models of musical expectancy, and can be used to generate musical sequences. One application tested the ability of a network to model schematic and veridical expectancies for musical chord sequences representative of Western music of the common practice era (Bharucha & P. Todd, 1991). Only short sequences (about 7 chords) were tested, but the network was able to learn some of the sequential regularities of Western harmony. P. Todd (1991) has used a similar network for melody learning, providing a test of the approach for longer sequences. This network was evaluated as a means of algorithmic composition. P. Todd (1991) found that the network was good at generating short musical lines, high in local structure, but lacking in global organization.

A difficulty with this approach is that the exponential recency gradient limits the ability of STM to adequately capture sequence context. STM loses information at a fixed rate, and this poses the challenge of maintaining STM traces long enough to contribute adequately to processing. Exponential trace memory is an example of an approach that considers time as a constraint on the maintenance of sequence history in STM. In attempts to make recurrent networks more sensitive to global structure, augmented versions of recurrent architectures have been proposed. Burr and Miyata (1993) and P. Todd (1991) have proposed that hierarchically cascaded recurrent networks might solve this problem, by chunking shorter subsequences to maintain memories more efficiently.

Mozer (1991; 1994) has applied a more powerful recurrent architecture to the problem of learning musical sequences in a network called CONCERT. As the name suggests, CONCERT was evaluated as a means of algorithmic composition. The CONCERT architecture is similar to Elman's (Elman, 1990) design. STM is implemented as a form of exponential trace, however, the content of memory is a powerful transformation of the input and state learned using the back-propagation through time (BPTT) algorithm (Rumelhart, Hinton & Williams, 1986). Thus, the structure of STM is adaptable and this network should in principle be capable of capturing arbitrary sequential and/or temporal relationships, including metrical structure, grouping structure, and even time-span reduction. In practice, Mozer (Mozer, 1994) found that the network did a good job of capturing local structure, but did not capture global structure well.

Mozer (1993; 1994) suggests that the inability of this architecture to capture global context is due to a limitation of the BPTT training algorithm. Mozer (1994) attempted to improve performance by providing the network with units that operated at different time constants, to provide the training algorithm with a greater amount of sequence history. Note that this approach also views time as a constraint on the maintenance of sequence history in STM. This manipulation improved performance somewhat, but it did not result in a significantly improved ability to capture global structure (Mozer, 1994). Mozer (1993) suggests that such results signal a basic difficulty in using sequential architectures to match transition probability distributions of very high order.

Another system based on exponential trace STM, using a different task and processing strategy, was Gjerdingen's (Gjerdingen, 1991; Gjerdingen, 1990) *L'ART pour l'art* network. *L'ART pour l'art* was designed to test the capabilities of a class of self-organizing networks based on adaptive resonance theory (Carpenter & Grossberg, 1987; Grossberg, 1976) in the musical domain. Short-term memory patterns were distributed representations of past input events, encoding features such as scale degree, contour, and inflection (Gjerdingen, 1991). Patterns in STM were categorized by a second level of units and stored in long-term memory as a pattern of synaptic strengths using a variant of Hebbian learning. Gjerdingen (Gjerdingen, 1991) tested the ability of ART 2 networks to make musically valid categorizations of the type of complex patterns that occur in passages of Mozart's six earliest compositions. With the simple exponential trace STM in place, the network tended to categorize incidental events as unique entities, preventing it from recognizing the types of underlying similarity that would allow it to create musically relevant generalizations (Gjerdingen, 1991).

Gjerdingen enhanced STM design by coding temporal relationships using five metrically oscillating levels of attention (Gjerdingen, 1991); events occurring at metrically important times were given increased activation in STM. This manipulation departs from the idea of exponential trace memory. Time was not viewed only as a constraint on the maintenance of sequence history in STM; relative time relationships were used to affect sequence processing. Gjerdingen (1989) compared the two STM strategies – exponential trace memory and metrically modulated exponential trace. The new network showed substantial improvement in the way it handled passing tones and other subsidiary events. The categorizations developed by the network were characterized as prototypes, or schemata, corresponding to musical concepts such as *galant*

cadence. The improved network made musically valid categorizations, although only for short segments of musical material.

Gjerdingen (1991) and Page (1993) report attempts to scale up these results to longer musical segments using hierarchically cascaded ART networks. These efforts are based on more sophisticated paradigms including ART 3 (Carpenter & Grossberg, 1990), masking fields (Cohen & Grossberg, 1987), and the SONNET 1 architecture (Nigrin, 1990), and have yielded promising results.

Wang and Arbib (Wang & Arbib, 1990; Wang & Arbib, 1993) have proposed another approach to modeling the recognition and production of complex sequences. Learning is based on a form of template matching using Hebbian learning. Wang and Arbib (Wang & Arbib, 1990; Wang & Arbib, 1993) showed that their recognition algorithm plus either exponential trace or interference-based (discussed below) STM models can learn to recognize any complex sequence, providing a solution to the problem of representational adequacy in STM. Their approach to encoding context is based on anticipation. Each recognition unit has an associated degree parameter. During learning, when a system's anticipated continuation of an input sequence is ambiguous (there is more than one possible continuation), the degree parameters of the active recognition units are incremented, so that they will detect longer contexts on the next training pass, and training continues until no ambiguities exist.

This approach has not yet been applied to music. However, the basic idea is similar to Kohonen's (Kohonen, 1984; Kohonen, et. al. 1991) dynamically expanding context (Wang & Arbib, 1993), which has been applied to musical sequence generation with interesting results. Mozer (Mozer, 1991) points out a potential problem with dynamically expanding context as a general approach to music, however. A particular note, i, can not be used to anticipate (or generate) a later note $i+n$, unless all intervening notes, $i+1$,..., $i+n-1$, are also considered. Thus, on presentation of the training sequences [A B C] and [A D C] the system would learn two unambiguous subsequences [B C] and [D C]. However, it would not make the generalization [A ? C], as might be appropriate in the musical domain. This analysis suggests that the dynamically expanding context approach is not appropriate for making generalizations concerning relative importance as, for example, the RAAM network did.

In summary, choice of STM model, processing, and training algorithm make important differences in how networks learn musical sequences. Models based on exponential trace STM designs can make musically valid generalizations, but have difficulty scaling up to the representation of longer musical sequences. Other approaches may recognize and generate individual sequences of arbitrary length and complexity (Wang & Arbib, 1990; Wang & Arbib, 1993; Kohonen, et. al. 1991), yet fail to generalize in musically appropriate ways. Alternative designs will be considered in the next section, however, some preliminary conclusions can be drawn at this point. In many of the approaches discussed above, time enters the picture only as a constraint on the maintenance of sequence history in STM. This may be a limiting feature: such approaches focus on sequence structure, minimizing or ignoring the importance of temporal structure, as defined above.

An exception was Gjerdingen's adaptation of the exponential trace design that allowed temporal structure to affect processing. The introduction of metrically oscillating levels of attention made *all the difference* in whether or not the system made the musical generalizations of interest. Another exception was the RAAM network that made even stronger assumptions about the effect of temporal structure on processing. The result was a system that represented long, complex musical sequences and made musically relevant generalizations.

5.1.2 Temporal Processing

This section examines how approaches to temporal sequence processing address the temporal structure of musical sequences: How does each architecture deal with time? There are three levels of assumptions that temporal sequence processing systems may make regarding time. These assumptions have to do with the meaning of 'time' in the model (the quantity represented by the parameter t in the model equations). First, time (t) may refer to real-time. *Real-time systems* must face the quantization problem head-on. This is the case in most speech processing systems, and some connectionist music processing systems. Second, time (t) may refer to idealized, or categorical, temporal durations as would be found in a musical score. *Relative-time systems* sidestep many difficult problems of temporal processing, and assume that the quantization problem has been solved (perhaps by preprocessing). This is the case in most connectionist music processing systems. *Serial-order systems* abstract away time all together, and deal only with the ordering information in a sequence of events.

5.1.2.1 Temporal Sequence Processing in Relative-Time

Most temporal sequence processing networks that have been applied in the musical domain are relative-time systems (e.g. Large, Palmer, & Pollack, in press; Mozer, 1991; P. Todd, 1991). Time is represented using categorical durations, as would be found in a musical score, making information about relative timing directly available to the network. Such systems do not address the quantization problem; they tacitly assume that this difficult problem is solved in preprocessing. There are several interesting questions that they may address, however. What effect can/should information about relative duration have on further processing of a musical sequence? Can the network learn to use temporal structure? Can the network be designed to exploit temporal structure?

In one such study, a simple recurrent network (Elman, 1990) was trained on a section of *The Blue Danube* Waltz (Stevens & Wiles, 1994). The musical sequence was input to the network as a sequence of events with pitch, accent, and duration properties. Thus, this relative-time network used a nonuniform sampling rate, representing duration as simply another event property. Stevens and Wiles (1994) gauged the performance of the network by comparing temporal and accent regularities extracted and represented in the network with the statistical properties of these components in the training composition. Expected frequencies of accent-duration pairs, such as quarter-note coupled with strong accent, were compared with actual frequency of occurrence in the composition. A canonical discriminant analysis of duration accent pairs by position in bar showed

that the hidden unit space was structured around inferred variables as well as around observable variables. This would suggest that the network may have learned something about the meter of the piece, but an analysis of whether or not meter was actually learned was not reported.

Gjerdingen (Gjerdingen, 1989) addressed the question of how prior knowledge of metrical structure may be used within a neural network for categorizing musical phrase types. Prior knowledge of metrical structure affected the way events were coded in STM. The network studied was a real-time/relative-time hybrid. First events were coded in STM using decaying activation based on event duration, and the network was trained at different rates of presentation. In addition temporal *relationships* were coded in STM using five metrically oscillating levels of attention (1989). Events occurring on *strong beats* (in the metrical structure) were given increased activation levels on entry into STM. Comparison of the two STM strategies – exponential trace vs. metrically modulated exponential trace – revealed that the addition of information about relative timing relationships (metrical structure) made *all the difference* in whether the network learned musically valid categorizations for the input patterns (Gjerdingen, 1989).

The RAAM networks reported in the previous chapter made use of temporal structure to produce representations that captured two kinds of sequence structure. First, the network represented the constituent structure of the musical sequences. It used knowledge of temporal organization (time-span segmentation) to adapt its processing strategy at each level, compressing and reconstructing groups of either two or three elements, to serve as an efficient encoder of predetermined structure. This chunk-and-recode strategy allowed the network to successfully represent long, complex musical sequences. Second, the reduced descriptions captured an important form of structural relationship among sequence elements, the relative importance of musical events. To accomplish this, the network used prior knowledge of metrical structure to learn stylistic regularities known to be systematically related to each element's metrical position.

The network learned relative importance because each position in its input buffer corresponded to a metrical grid location, and the network used a dedicated set of weights for each position. This strategy made the network sensitive to relative timing relationships in a unique way. Consider the coding of the two sequences X1= [A^{360} B^{120}] and X2 = [A^{720} B^{240}] as coded by the second RAAM network (see Section 4.4, and Figure 19). The temporal relationships correspond respectively to the relative timing relationships of dotted eighth note followed by sixteenth note (the initial rhythmic figure of *Mary had a little lamb*), and a dotted quarter note followed by an eighth note (the same rhythmic figure slowed by one-half). The network input for the first figure is the tree (A (NULL B)) and for the second figure is the same tree (A (NULL B)). It produces the *same code* for both figures because the *same weights* are used, corresponding to the same *relative* metrical grid locations. The (qualitative) difference in rate is only coded in relationship to the rest of the melody. A similar argument can be made for the fully balanced tree RAAM, but in that case the codes are just similar, not exactly the same. RAAM learned metrical accent and time-span reduction because it was able to exploit to relative time relationships.

5.1.2.2 Temporal Sequence Processing in Real-Time

In a relative-time system time delays are easy to think about. Consider the RAAM network at the moment it recodes two consecutive quarter notes. At the point in time when the network "fires," input consists of a signal corresponding to the current pitch event, and a signal corresponding to a previous pitch event delayed by a fixed, discrete amount of time. Specification of such a system is simplified because 'quarter note' is a relative duration. In a real-time system, however, the situation is more difficult. Consider the same situation, but in real time, with a sampling rate of 1 *ms*, and an average quarter note duration of 480*ms*. With a fixed, discrete time delay (e.g. Lang, Waibel, & Hinton, 1990), if the inter-onset-interval (IOI) is equal to 479*ms*, when the new event enters memory and the network fires, the previous input will not be available. This is an example of the quantization, or time-warping problem.

Real-time systems must deal with the time-warping problem directly. One way to deal with this problem is to not simply delay the signal corresponding to a previous event, but to also convolve it with a broadening function (Tank & Hopfield, 1987). With respect to the previous example, this would make the signal available to processing at a range of times around 480*ms*, with the strongest response at precisely 480*ms*. Thus some deviation in an average period of 480*ms* can be tolerated and the system will still behave adequately. As discussed by de Vries and Principe (de Vries & Principe, 1992) and Mozer (Mozer, 1993) temporal convolution represents a general delay mechanism. From this point of view, for example, the impulse response of simple delay line is a delayed unit sample. The impulse response of an exponential trace memory is $(1-\mu_i)\,\mu_i^t$ where $-1 \le t \le 1$, an exponentially decaying function.

Convolution with a broadening function has been applied with some success in speech recognition (e.g. Tank & Hopfield, 1987). deVries and Principe (1992) characterize such memories using two parameters, depth (delay time) and resolution (amount of broadening). They propose the g*amma* neural model that provides efficiency benefits compared with gaussian delay kernels, and includes both discrete delay kernels and exponential kernels as special cases. Several researchers have also used periodic memory kernels, and have addressed the problem of learning the parameters of memory kernels during batch training (Unnikrishnan, Hopfield & Tank, 1991; Bodenhausen & Waibel, 1991; de Vries & Principe, 1992; Principe, de Vries, & de Oliveira, 1993). The difficulty with these approaches is that, whether delays are hardwired or learned during batch training, they remain fixed during sequence processing. Thus such memories may be expected to show difficulties in processing musical sequences.

Given this new perspective on the temporal processing properties of delay kernels, one might expect an exponential trace memory to show some robustness in the face of timing deviation. For small changes in presentation rate, small changes in processing results may be expected. McGraw, Montante, and Chalmers (1991) tested this intuition in the musical case. They attempted to train various recurrent networks as simple "beat detectors," but found that a network trained to output beats to one melody at three different tempos did not correctly respond to the same melody played at a fourth, intermediate tempo. Thus in practice, recurrent networks have been shown to generalize poorly to novel presentation rates, relying upon absolute rate information to process temporal patterns. According to results such as this, one may expect that a network trained to

recognize a melody played at 80 beats per minute, for example, would not recognize the same melody played at 90 beats per minute.

Cottrell, Nguyen, and Tsung (Cottrell, Nguyen, & Tsung, 1993) have attempted to solve this problem with a strategy for rate invariant sequence recognition in simple recurrent networks. They first trained a recurrent network to predict a target input signal presented at some "normal" rate. A typical recurrent network would track the target signal at this rate, but would lose the signal at other rates. Cottrell et. al. augmented their network to control its own processing rate by adapting time constants and processing delays. Using prediction error, the recurrent network adapted its processing rate to match the rate of the current signal, much like a phase-locked loop varies its internal frequency to match the phase of an incoming signal. This approach worked well in the test domains in which it was applied. The drawback of this approach is that it applies only to learned sequences.

Others have attempted to solve this problem in a recurrent network trained with the real-time-recurrent learning (RTRL) algorithm (Anderson & Port, 1990; Cummins, Port, McAuley, & Anderson 1993). The representations developed by the network were viewed as trajectories through activation space sculpted by a chain of stable attractors. The training procedure moved the locations of the attractors for pattern elements and the locations of 'recognition regions' so that learned patterns could be differentiated from distractors. After training the network reliably recognized learned patterns at faster and slower rates of presentation, but also at irregularly altered rates. Wang and Arbib (1993) have achieved a more general result using an STM model based on interference: a unit's input activation does not decay with time, but with the number of other items currently held in short term memory. In this system the network is sensitive only to serial order information. Because both networks are insensitive to temporal information, they cannot differentiate between sequences based on relative timing information, and time is treated as *serial order*. Serial order systems are of limited utility in the musical domain.

In summary, temporal sequence processing networks can deal with time in one of three ways: as real-time, as relative-time, or as serial order. Real-time networks must deal with the quantization problem, the problem of processing temporal sequences in a rate-invariant, relative-time sensitive way. State-of-the-art real-time systems use STM structures based on time-delays and broadening functions learned during batch training; during on-line sequence processing STM structures remain fixed. Relative-time networks rely on the tacit assumption that the quantization problem can be solved by other means. For example, a separate system may adjust processing rate according to the rate of the incoming sequence for learned sequences (Cottrell, Nguyen, & Tsung, 1993). Relative-time systems are useful in investigating how neural networks can learn and/or make use of patterns of relative timing. For instance, recurrent networks may be able to learn some metrical relationships (Stevens & Wiles, 1994). Other networks, such as Gjerdingen's (1989) and the RAAM network of the previous chapter exploit knowledge of metrical structure. Networks that make use of metrical structure to organize short term memory traces have shown the greatest ability to make musically and psychologically valid generalizations. These results are consistent with the psychological results cited in Section 2.3 and with dynamic attending theory (Jones, 1976; Jones & Boltz, 1989).

5.1.2.3 From Real-Time to Relative Time

What would be most desirable is a system that works in real-time, and is sensitive to both simple relative timing relationships and full blown metrical structures, as humans are. Such a system would be rate-invariant, relative-time sensitive, and would make musical generalizations, much as humans do. STM kernels with fixed depth and resolution are a step in the right direction, but ultimately they will not do the job. As shown in Chapter III, expressive timing deviations in musical performance and improvisation result in timing structures that are far too flexible to yield to temporally static memory structures. On-line adjustment of memory parameters (Cottrell, Nguyen, & Tsung, 1993) represents another step in the right direction, but currently proposed strategies apply only to learned sequences. What is needed is a strategy by which unlearned sequences can also be processed efficiently.

Let us postulate an STM kernel function with a periodic impulse response. Three parameters characterize this function: *period*, *resolution,* and *decay.* Let us further suppose that this impulse response function is somehow able to automatically adjust its parameters so that points of maximum output correspond to beats at some level in a metrical structure grid; the parameters of this function automatically adjust on-line as performance tempo increases or decreases. Because this hypothetical function has a period and phase corresponding to a level of beats in a metrical structure grid, it will do a better job than a fixed memory kernel of dealing with time-warping, and it can adjust parameters without memorizing the sequence in advance. Several such functions, each corresponding to a level of beats in a metrical structure grid, would embody knowledge of relative time relationships and temporal structures. Based on such a mechanism one could expect results as good as Gjerdingen's (1989), the RAAM model's, or perhaps even better, in a *real-time* (as defined above) network. The next three chapters propose a way of doing precisely this: a way of getting *real-time* networks to perform as well as *relative-time* networks.

Jones (1976), Jones & Boltz (1989), Gjerdingen (1989) and others have suggested the presence in the brain of oscillatory assemblages of neurons that can entrain themselves to periodic signals such as those found in musical rhythms. This suggestion presents one possibility for creating an adaptive memory kernel. An oscillator (to be defined) entrains to a pseudo-periodic component of a perceived input rhythm. The oscillator generates signals (beats) corresponding to the phase, period, and variability of the rhythmic component that it tracks, so that period, resolution, and decay of *any* memory kernel function can be adjusted on-line. The unit ignores the content of the sequence (events), dealing only with the rhythm (the pattern *of* time). Therefore, the strategy works both for learned and unlearned sequences. I will not implement any memory kernel models; rather I propose a method that will enable a variety of possible STM designs that automatically adapt to temporally structured input signals. I make the strong claim only that this approach will work for the particular types of temporal structure found in music (metrical structure). However, this approach may work for other domains as well. In chapter IX, I address the issue of speech recognition more directly.

Mine is not the first attempt to entrain a signal generator to complex rhythmic sequence. This problem has been studied in the music-processing literature under the name *beat-tracking*. The remainder of this chapter will focus on previous systems proposed for dealing exclusively with

patterns *of* time (rhythms), focusing on the computation of temporal structure in music. The following three chapters will propose an entrainment model.

5.2 The Computation of Temporal Structure

Temporal structure plays an important role in the organization of human perception, however, mechanisms for the perception of temporal structure are still poorly understood. As noted in Chapter 3, temporal deviation in musical performance limits the utility of straightforward applications of well-understood signal processing methods such as Fourier analysis and auto-correlation. Symbolic approaches relying on the parsing of temporal patterns have been proposed (e.g. Jackendoff, 1992; Longuet-Higgins & Lee, 1982; Scarborough, Miller & Jones, 1992), but again due to temporal deviation, these methods do not handle musical performance. Entrainment, or synchronization to a perceived beat, may provide some answers. I will discuss several models related to the perception of metrical structure, illustrating the problems entailed by the design of entrainment mechanisms for the perception of complex musical rhythms.

5.2.1 Quantization and Time-Warping

Well-known signal processing and information processing techniques may be applied to certain rhythms to recover metrical structure if presented with a stationary signal as input. For this reason, quite a bit of work has been done with preprocessing rhythmic signals, cleaning up messy timing data, so that techniques designed for stationary input may be applied. Such approaches are usually referred to as *quantization* or *time-warping* approaches. The most straightforward approach to quantization is to do it by hand. For example, to prepare musical sequences for input to the neural network model described in Chapter IV, a commercial music software package was used that allows the user to perform quantization through an interactive process. The problem with the user-interaction approach in this context is obvious, the user becomes a *homunculus* in the theory of temporal sequence processing.

Desain and Honing (1991) developed a connectionist quantizer to automatically "clean up" messy timing data in music so that the metrical structure may be inferred. From their point of view, the relevant task is one of inferring from the inter-event time intervals in the signal the ideal, or intended, inter-event intervals. Using a constraint-relaxation technique the quantizer works on a window of intervals to adjust inter-event durations so that every pair of durations in the window is adjusted toward an integer ratio, if it is already close to one. The main advantage of this technique is the relatively weak assumptions made about the nature of the input rhythm (integer time ratios are to be preserved). A disadvantage is that even these relatively weak assumptions may be too strong to represent a general solution. Divisive rhythms (for example, a group of two followed by a group of three) are problematic for this approach (Desain & Honing, 1991). Although they do not specifically address the issue, this approach would have similar problems with polyrhythms common in music. A second disadvantage is that the algorithm is inefficient, and thus of questionable utility for real-time analysis. A further disadvantage is that it works on a fixed-size input window, whose size must be adjusted depending upon the nature of the input.

5.2.2 Structural Analysis of Rhythmic Signals

One approach to the structural analysis of rhythmic signals involves the adaptation of signal processing methods developed for stationary signals to the processing of nonstationary input. Such approaches generally involve windowing the input, assuming that the signal is locally stationary within the input window. Such approaches have the additional advantage of increased efficiency over considering the entire input signal at once. An example of such an approach is narrowed auto-correlation (Brown, 1992; 1993; Brown & Puckette, 1989). Narrowed auto-correlation has been successfully applied to the problem of determining the meter of musical scores (Brown, 1993). This approach is successful because music is often composed in such a way that more events occur at strong metrical locations (Palmer & Krumhansl, 1990). Musical scores, however, do not contain temporal deviations. Brown (1993) has reported preliminary results in applying this method to a segment of one performance. Further study is required to determine the general applicability of this approach.

One method of structural analysis proposed by several researchers (e.g. Longuet-Higgins & Lee, 1982, Jackendoff, 1992) is to parse a rhythm according to a context-free grammar. A set of rules is constructed that describes allowable temporal structures, and well-known algorithms use the rules to identify the structure in the input. An advantage of this approach is that the analysis implicitly performs structure recognition. A disadvantage is that it applies only to musical scores, not to performed music. Also real-time parsing may require either backtracking or simultaneous consideration of multiple alternative structures, and both strategies hamper efficient processing. Finally, context-free parsing assumes nested hierarchical structures, and thus cannot efficiently account for the perception of polyrhythmic structure.

Scarborough, Miller, & Jones (1992) have described a model of meter perception called BeatNet, based on a parallel constraint satisfaction paradigm. Conceptually, the BeatNet network is a one-dimensional array of idealized low-frequency oscillators with different beat-periods that operate to align their output "ticks" with event onsets. Output of the system is a metrical grid of the style proposed by Lerdahl and Jackendoff (Lerdahl, & Jackendoff, 1983). A metrical structure emerges from local interactions between oscillators, rather than from the global effect of rule-based analysis. An advantage of this approach is that it handles the problem of metrical preferences through real time processing constraints, rather than by global evaluation of alternative constructs. A disadvantage of this method is that it does not handle performance timing, because phase and period of each oscillator is fixed.

5.2.3 Dynamic Processing of Input Rhythms

In some situations it is important to be able to process signals on-line, with only local information. Such approaches are sometimes called beat-tracking approaches. The idea of beat-tracking is to synchronize an internal signal generator (generating beats) with a component periodicity of the input rhythm. Several approaches to the problem of beat-tracking have been proposed (e.g. Allen & Dannenberg, 1989; Dannenberg, 1984; Dannenberg & Mont-Reynaud, 1987; Longuet-Higgins, 1987; Rosenthal, 1992; Vercoe & Puckette, 1985). The length of the beat-

period is adjusted throughout the rhythm as the performer speeds up or slows down. Thus beat-tracking attempts to deal with nonstationary input signals. For example, Dannenberg & Mont-Reynaud (1987) describe a history mechanism that uses a weighted average of previous perceived tempos to compute current perceived tempo. Allen & Dannenberg (1989) use a state description that includes phase and period, and real-time beam search to allow the beat-tracker to consider several possible states at once. A potential problem with each of these approaches is that symbolic implementation forces the algorithms to make discrete choices, requiring explicit (sometimes simultaneous) consideration of multiple alternatives. Such approaches may underestimate the dynamic complexity of the beat-tracking task.

Longuet-Higgins (1987) proposed a hybrid method that combines beat-tracking with metrical structure parsing to perform structural analysis of nonstationary input signals. The program uses a static tolerance window, within which it will treat any event as "on the beat." Events which fall outside the window are interpreted as subdividing the beat into groups of either two or three. This approach considers the tracking of individual levels of beats with in the larger context of meter perception, and may result in improved performance. A disadvantage of this approach is that it suffers from potential inefficiencies of context-free parsing, including difficulty in accounting for polyrhythmic structures.

Several connectionists have proposed entrainment mechanisms for beat and meter perception (Large & Kolen, 1993; in press; McAuley, 1993; 1994; Page, 1993). Page (1993) proposes that a neural entrainment mechanism should operate analogously to a *phase-locked loop*, an electronic circuit commonly used in communications applications. Page (1993) recruits a network of connectionist units into a neural implementation of a standard phase-locked loop. The heart of the network is a gated pacemaker circuit (Carpenter & Grossberg, 1983). Page implements a Type-II phase detector and a low-pass filter using networks of connectionist units, to provide an error signal that controls adjustments to phase and period in the gated pacemaker. However, there are several problems associated with this approach, detailed in Page's (1993) simulations. Most importantly, Page's (1993) design assumes that the input signal is periodic. This assumption places limitations on the circuit's ability to deal with the complex rhythmic structures of music. Because the phase-locked loop reacts to every input event, it cannot extract a "component periodicity" from a complex rhythmic pattern. Page deals with this problem by assuming that relevant periodicities are unambiguously marked in the signal via phenomenal accent information. In music, however, phenomenal accent information is often missing, ambiguous, or even misleading (e.g. syncopation).

An important research problem for entrainment approaches is to find an appropriate type of oscillator for modeling musical beat. To illustrate the relevant issues, consider a simple model that has been used as a model of single cell oscillation in the nervous system, the integrate-and-fire oscillator (Glass & Mackey, 1988; Winfree, 1980). The simplest formulation of the integrate-and-fire model is shown in Figure 21. Activation increases (linearly) to a threshold, the unit fires, resets its activation to zero, and the process begins again. As shown in Figure 21A, the unit spontaneously oscillates with a period determined by the slope of the activation function and the height of the threshold. Figure 21B shows the unit phase-locking to a discrete periodic stimulus. Each discrete

stimulus event temporarily lowers the unit's threshold so that the oscillator may fire and reset earlier than would otherwise be the case. Figure 21B also illustrates one problem with phase-locking oscillators as models of musical beat. When the stimulus ceases, or when an onset is missing, the oscillator immediately reverts to its original period, as though no stimulus had ever been present. In other words, the oscillator has no memory of the previous rhythmic context. Torras (1985) proposed a scheme for frequency tracking in a different integrate-and-fire model. In this formulation, an integrate-and-fire oscillator can adapt to the frequency of a stimulus by adapting its threshold. This situation is shown for the simpler model in Figure 21C. McAuley (1993) proposed that a Kohonen map of Torras oscillators could memorize, categorize, and reproduce musical rhythms.

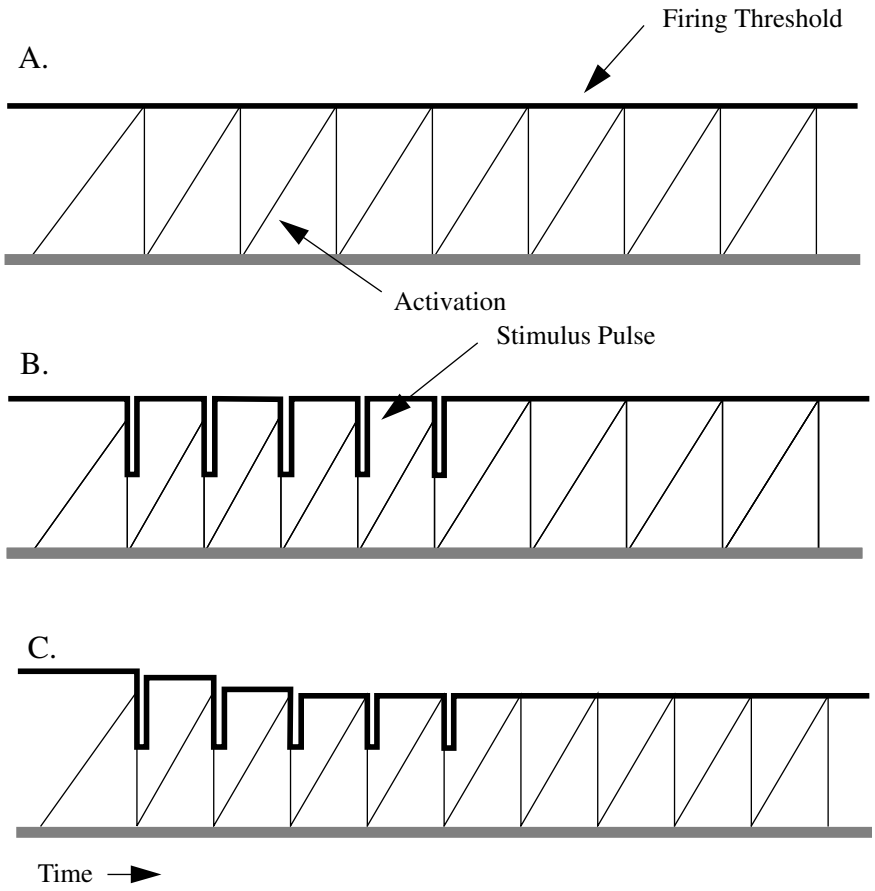


Figure 21: A periodic signal and the response of an integrate-and-fire oscillator.

Integrate-and-fire units have their own set of problems in the domain of meter perception. For example, the discontinuity in the activation function constrains the oscillator to adjust its period only by speeding up (McAuley, 1994). Large and Kolen proposed a continuous model to avoid this problem and the problems exhibited by phase-locked loop models (Large & Kolen, 1993; in press). The model presented in the following chapters is an extension of that proposal.

McAuley (1994) recently compared the performance of four different oscillatory units including two integrate-and-fire models, Large and Kolen's original model (Large & Kolen, 1993), and a simplification of that model. McAuley (1994) prefers the simpler model. However, this simplification creates problems similar to those found in phase-locked loop models; both require strong assumptions about phenomenal accentuation to display appropriate behavior.

In summary, the difficulty of identifying metrical structure in real-world signals arises from rhythmic complexity (missing and extraneous events), timing deviations, and structural complexity (polyrhythms). Without these difficulties, metrical analysis could be performed by any of the methods described above. A successful mechanism must be able to "pick" pseudo-periodic components out of a complex rhythmic pattern in spite of missing, ambiguous, or misleading information, and combine these components into complex structures. The following chapters propose such a mechanism.

CHAPTER VI

SYNCHRONIZATION TO COMPLEX SIGNALS

The perception of beat and meter is a fundamental cognitive/perceptual ability. In humans, this ability enables apparently simple behaviors including tapping along with a tune, and very complex behaviors including the ability of skilled musicians to coordinate intricate motor programs with perceived musical rhythms. It may also enable more general abilities such as rate-invariant temporal sequence recognition. Synchronization with isochronous input signals is relatively easy to achieve, however synchronization with complex signals can be quite difficult. This chapter presents a model of synchronization that is appropriate for complex, temporally structured signals, and is motivated from the perspective of music perception and cognition. For reasons cited above, however, it may have wider applicability.

The difficulty of identifying temporal structure in complex signals such as music arises from three sources. The first source of difficulty is the presence of *systematic timing deviations*. In music, performers use temporal deviation, or rubato, to communicate musical intentions. Such systematic deviations produce nonstationary input signals, limiting the usefulness of analytical techniques designed for stationary signals, such as Fourier analysis. The second source of difficulty is *rhythmic complexity.* In music, rhythmic complexity refers to factors including amount of syncopation and number of different duration values present in a rhythm. In other words, the "periodic components" of rhythms that correspond to beats are not really periodic. Even in ideally timed rhythms there are missing events and extraneous events. The third source of difficulty is *structural complexity.* The existence of polyrhythmic structure, in particular, limits the usefulness of approaches that assume strict temporal nestings, such as context-free parsing. Without these difficulties, metrical analysis could be performed by more traditional approaches. A successful mechanism must be able to "pick" pseudo-periodic components out of complex rhythmic patterns in spite of missing, ambiguous, or misleading information, and combine these components into complex structures.

This chapter describes a mechanism for beat perception in complex, metrically structured rhythms that addresses these difficulties. The mechanism works on-line with local information. It possesses a memory for recent events, displays expectations for upcoming events, and can handle missing events at those times. The mechanism can also ignore events that should not affect its behavior. These properties are achieved by synchronizing, or entraining, an oscillator to an incoming signal. The oscillator generates output pulses with a given phase, period, and width. When an event occurs during an output pulse, the oscillator will adjust its phase and/or period to align the output pulse with the input event. The output pulse prevents intervening events from

distracting the oscillator, because the oscillator will ignore events that fall outside its output pulses. A system of oscillators can be used to recover the metrical structure of input signals, with different oscillators generating different levels of beats.

6.1 Definitions

Let the term *rhythm* refer to a time-series of events. Some rhythms display a form of temporal organization called *metrical structure* (Essens & Povel, 1985; Lerdahl, & Jackendoff, 1983). A metrically structured rhythm can be defined as a rhythm generated by a coordinated set of oscillatory event generators. The phases and periods of individual event generators may vary, however certain temporal relationships between event generators remain invariant over time. A metrical structure is a set of *perceived* relationships between (assumed) event generators. The problem of metrical structure is to identify and track the components of a rhythmic signal corresponding to individual event generators, and to identify and track relationships (relative phases and periods) of the event generators, in real time. One way to think of this perceptual task is as the recovery of a (motor) program structure that would recreate the rhythm – a program built of oscillatory event generators.

According to this perspective, metrically structured rhythms are composed of several not-quite-periodic components, each corresponding to the output of an individual event generator. These not-quite-periodic components are called *pseudo-periodic event trains*. Each *event train* corresponds to the output of one event generator over time. An event train is *pseudo-periodic* because it arises from a nonstationary source – the phase and period of its generator may change. When one taps one's foot to music, one is tracking a single pseudo-periodic event train in the musical rhythm. The brain is synchronizing an *internal event generator* to a perceived external event generator. Internally generated events correspond to a level of perceived *beats,* as defined in the Chapter II. Metrical structures can be defined as relative phase and period relationships between these internally generated levels of beats.

6.1.1 Synchronization and Entrainment

Events can be represented in an input signal as discrete impulses, $s(t)$. Figure 22 shows a series of impulses, corresponding to note events in an improvised melody, collected on a computer-monitored piano (see Chapter III). In Figure 22, $s(t) = 1$ when an event occurs, and $s(t) = 0$ at other times. In general, $s(t)$ may take on any value, and it is sometimes useful to let the value of $s(t)$ carry information such as amplitude.

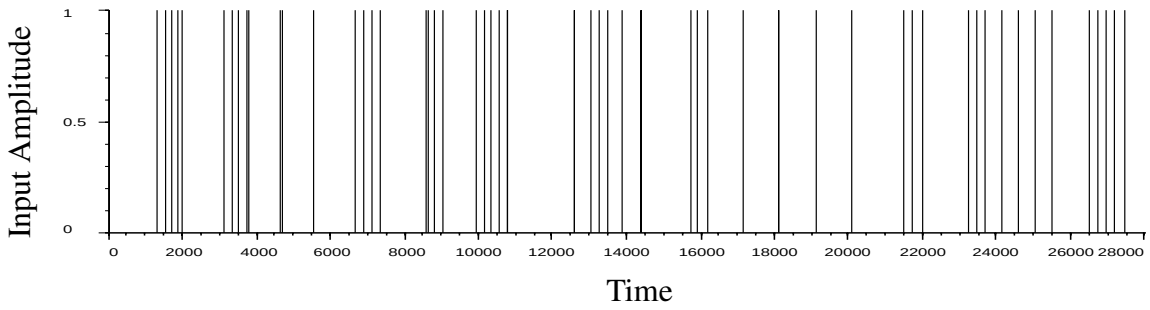


Figure 22: An input signal to the oscillator model

Beats (internally generated events) can be modeled using an oscillator that generates events at some specific point in its cycle. Oscillations repeat after some specific interval of time, called the *period*, p, of the oscillation. *Phase* at time $0 < t < p$ can be defined as $\phi(t) = t/p$. According to this definition, phase lies between 0 and 1. Two oscillations are *synchronized* when they regularly come into phase, or begin their cycles together. A process by which two or more oscillations achieve synchronization is called *entrainment*. Entrainment occurs because a *coupling* between two or more oscillations causes them to synchronize. Coupling allows a signal (the driver) to perturb an oscillator (the driven) by altering its phase, its period, or both.

The approach is as follows. The rhythmic input signal serves as a *driver*, and impulses in the signal perturb both the phase and the period of a *driven oscillator*, causing changes to the oscillator's behavior. The oscillator adjusts its phase and period only at certain times during its cycle, isolating a single event train in the incoming rhythm. The oscillator generates events that correspond to beats.

6.2 The Oscillator Model

6.2.1 Output Events

The oscillator generates events periodically, and when driven by an input signal adjusts its phase and period so output events track the phase and period of a single pseudo-periodic event train in the rhythm. Two kinds of output events can be defined, discrete and continuous, and these are useful for different purposes. Discrete events are impulses, and can be generated as follows. Let t be time and t_x be the time of next expected event. When $t = t_x$ expected event time t_x is reset in anticipation of the next event, $t_x \leftarrow t_x + p$. At this time the oscillator generates a discrete event. These discrete events correspond to the music-theoretic notion of beat described in Chapter II. The continuous events are called *output pulses*. An output pulse is like a beat except that is has a width, an extent in time. Let the phase of the signal generator at time t, be defined as:

$$\phi(t) = \frac{t - t_x}{p}. \qquad \text{(Eqn 1)}$$

An output pulse may then be defined as:

$$o(t) = 1 + \tanh\gamma(\cos 2\pi\phi(t) - 1) \qquad \text{(Eqn 2)}$$

where γ is a parameter called *output gain*.

Figure 2 shows an output pulse, in the absence of input, as a function of time. Amplitude is maximum ($o(t) = 1$) at the beginning of each cycle (i.e. when $t = t_x$), quickly falls to zero for the body of the cycle, then begins to rise again to a maximum as the cycle comes to a close. Amplitude is only non-zero for a relatively small portion of the cycle. The output pulse defines a temporal receptive field for the oscillator, a region of temporal "expectancy". The oscillator entrains to the signal by adjusting its phase and period only in response to signal impulses that occur within this receptive field; it ignores impulses that occur outside of this field (when $o(t) \approx 0$). This allows the oscillator to identify and track a single event train in a complex signal, while ignoring irrelevant information.

The parameter γ, the output gain, determines the size of the receptive field. When γ is small, (Figure 2A), the receptive field is wide and the oscillator will tolerate a relatively large amount of variability in the input signal. When γ is large, (Figure 2B), the region is narrow and the unit will tolerate relatively little variability in the signal.

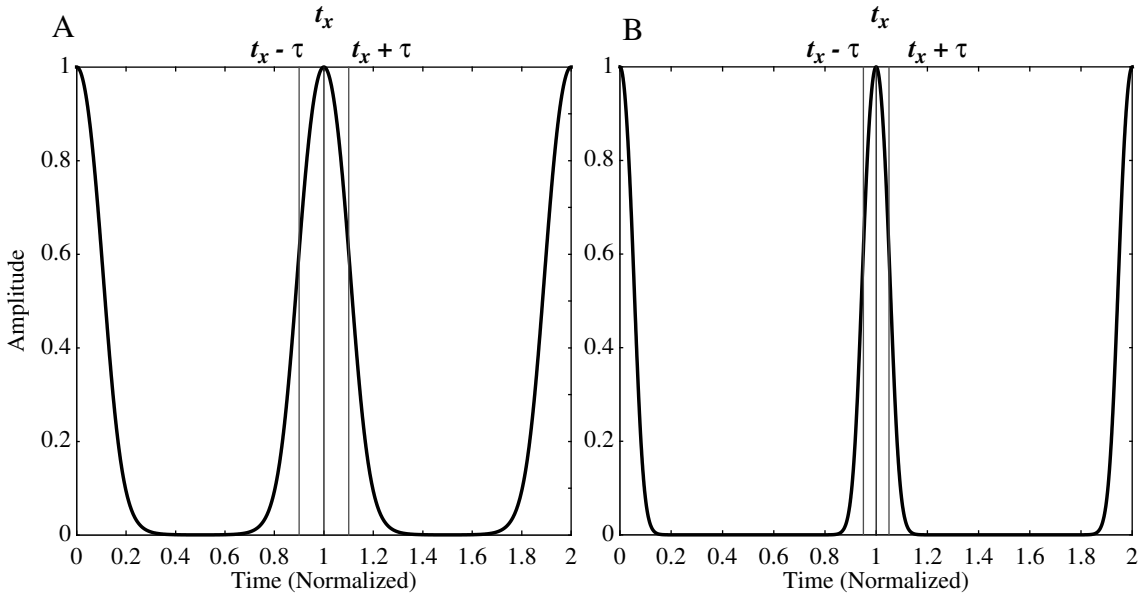


Figure 23: Output pulses (temporal receptive fields) for two different values of τ. (A) $\tau = 0.10$, (B) $\tau = 0.05$. τ measures the width of the temporal receptive field.

In order to deal with nonstationary signals efficiently, the oscillator adjusts the size of its receptive field as it entrains to the signal (described below). Once the oscillator has entrained, the value of γ acts as a measure of the variability in the target event train. Because this measure is somewhat difficult to interpret, I introduce a second, related measure, called τ. τ measures the width of the temporal receptive field as the distance between the points of inflection on the curve (corresponding to relative extrema in the first derivative, and zero-crossings in the second derivative). τ is related to γ as follows:

$$\gamma = \frac{\omega}{\cos 2\pi\tau - 1}. \qquad \text{(Eqn 3)}$$

In this equation, ω is a constant that can be determined to be $\omega = -0.416$. In what follows, I refer to τ rather than to γ.

6.2.2 Phase-tracking and Period-tracking

In order to track an input signal, the oscillator must adjust its phase and period in response to input impulses. To accomplish this, two types of behavior must be specified, phase-tracking and period-tracking. These behaviors can be specified by formulating delta rules that describe the amount of adjustment to the phase and/or period in response to an impulse. Phase-tracking behavior can be implemented with the following rule:

$$\Delta t_x = \eta_1 s(t) \frac{p}{2\pi} \operatorname{sech}^2 \gamma (\cos 2\pi\phi(t) - 1) \sin 2\pi\phi(t), \qquad \text{(Eqn 4)}$$

where η_1 is a parameter called *coupling strength*. Figure 24 shows the shape of this curve, summarizing the effect of the delta rule in response to an input impulse. This curve is related to the first derivative of Equation 2 with respect to time. The relative extrema correspond to points of inflection on the output pulse curve. τ measures the amount of deviation from t_x (as a percentage of the current period) that the oscillator will tolerate, and still adjust its phase and period to efficiently track an event train. The rule can also be thought of as a modified gradient descent rule, minimizing an error function that describes the difference between when impulses are expected and when they actually occur (Large & Kolen, in press). The presence of $s(t)$ in this formula ensures that adjustments to phase will occur only when a signal impulse is present ($s(t) > 0$). An impulse that occurs within the oscillator's temporal receptive field, but before t_x causes a negative phase shift, because $\Delta t_x < 0$. An impulse after t_x causes a positive phase shift, because $\Delta t_x > 0$. Thus, this delta-rule provides a non-linear coupling term implementing phase-coupled entrainment. Phase-tracking is most efficient when $|t_x - t| \leq \tau$. When, $|t_x - t| > \tau$ the adjustment to t_x is less than would be necessary to phase-lock the output pulses to the signal impulses. If $|t_x - t|$ is large enough, the unit will ignore the impulse. If $|t_x - t| > \tau$ for several cycles, the oscillator will lose the signal.

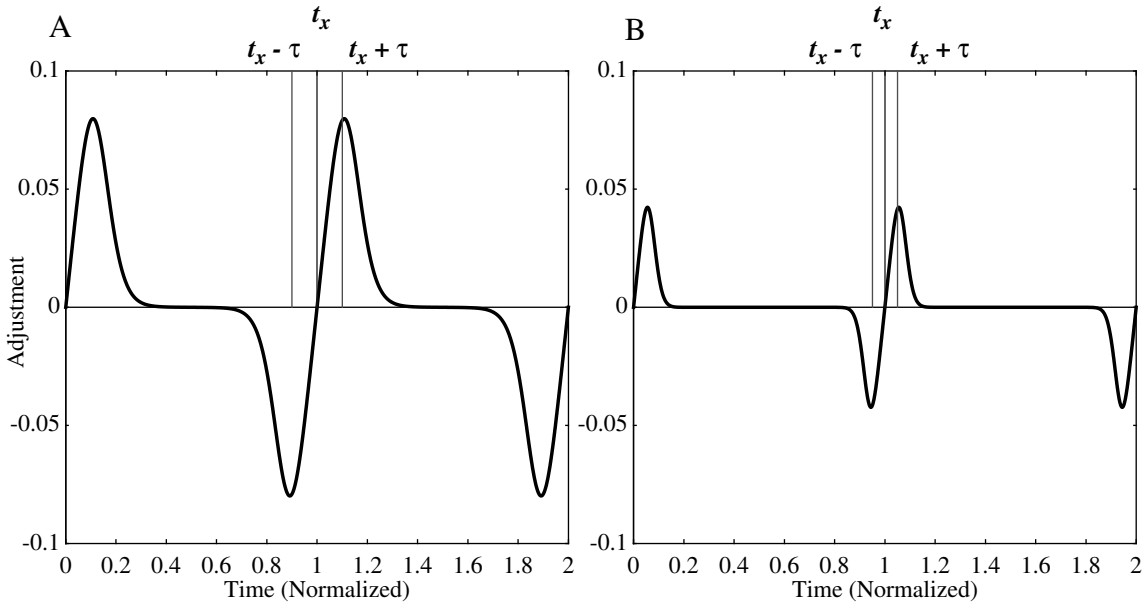


Figure 24: The effect delta rules for phase and period given in Equations (3) and (4) for two different values of τ, (A) $\tau = 0.10$, (B) $\tau = 0.05$. This figure illustrates how τ gives the amount of variability that the unit will tolerate input signal.

Period-tracking behavior may be achieved by noting that the difference between expected and observed period is the same as the difference between expected and observed impulse times (assuming the oscillator was in phase in the last cycle). Therefore, the same delta rule can be used, with the introduction of a new coupling term, η_2:

$$\Delta p = \eta_2 s(t) \frac{p}{2\pi} \operatorname{sech}^2 \gamma (\cos 2\pi\phi(t) - 1) \sin 2\pi\phi(t). \qquad \text{(Eqn 5)}$$

The use of a separate coupling term allows independent adjustment of the phase- and period-tracking behaviors. Aside from this difference, the rules are identical. Again, the presence of $s(t)$ in this formula ensures that the delta rule will have a non-zero value only when a signal impulse occurs. Figure 24 also shows the shape of this curve, summarizing the effect of the delta rule in relation to the output pulse. An impulse that occurs within the oscillator's receptive field, but before t_x causes a negative adjustment to period, because $\Delta p < 0$. An impulse after t_x causes a positive adjustment to period, because $\Delta p > 0$. Thus, this delta rule provides a non-linear coupling term implementing period-coupled entrainment. As above, period-tracking is most efficient when $|t_x - t| \leq \tau$, and if $|t_x - t|$ is large enough, the oscillator will ignore the impulse.

6.2.3 Tracking Variability

It is useful for the oscillator to adjust the size of its temporal receptive field. This allows the unit to adapt to the amount of variability in the input signal. To do this, it is necessary to create one more delta rule, a rule that adjusts τ. By adjusting τ the oscillator effectively estimates variability in the phase and period of the target event train. Thus it can adapt its temporal receptive field to efficiently track different types of signals. To accomplish this behavior, τ is limited to a fixed range between τ_{min} and τ_{max} by introducing the control parameter Ω, which is related to τ according to the following equation:

$$\tau = \tau_{max} + 0.5\,(\tau_{min} - \tau_{max})\;(1 + \tanh\Omega)\;. \qquad \text{(Eqn 6)}$$

Figure 25 shows this relationship.

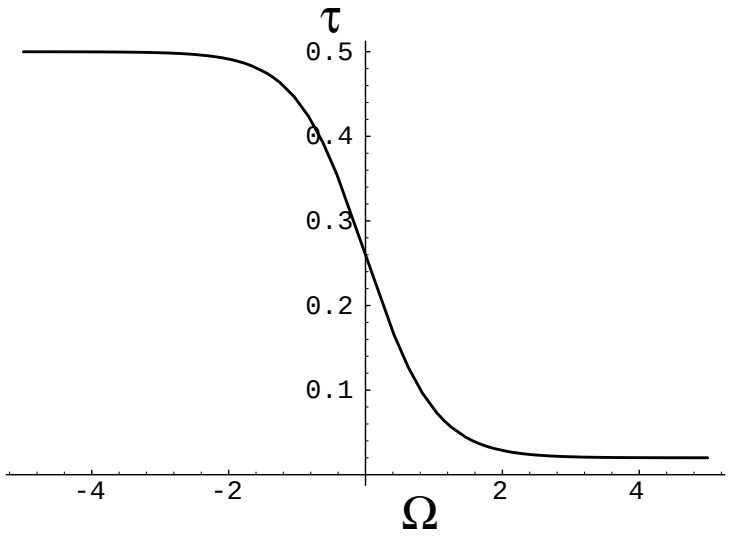


Figure 25: The relationship between Ω and τ, according to Equation 6.

A delta rule can then be defined to adjust Ω:

$$\Delta\Omega = \eta_3 s(t)\,\mathrm{sech}^2\gamma\,(\cos 2\pi\phi(t) - 1)\;(\cos 2\pi\phi(t) + 2\gamma\,(o(t) - 1)\,\sin^2 2\pi\phi(t))\,. \qquad \text{(Eqn 7)}$$

This delta rule is related to the second derivative of Equation 2 with respect to time. Figure 26 shows the shape of this curve, summarizing the effect of the rule. The zero-crossings correspond to the relative extrema of the phase- and period-tracking delta rules, giving this rule the power to adapt the oscillator's tracking behavior according to the variability of the input signal. An impulse that occurs when $|t_x - t| < \tau$ causes the oscillator's temporal receptive field to shrink, because it is doing a good job of predicting the input. An impulse that occurs just outside this region will cause

the temporal receptive field to grow, because the oscillator came close to predicting the impulse, but is attempting too precise a prediction. If $|t_x - t|$ is large enough, the this rule will ignore the impulse. Because impulses may not occur in every cycle, Ω decays toward 0 each time the unit generates an output event. Thus, if there is no event in the oscillator's current cycle, τ will increase, widening the temporal receptive field.

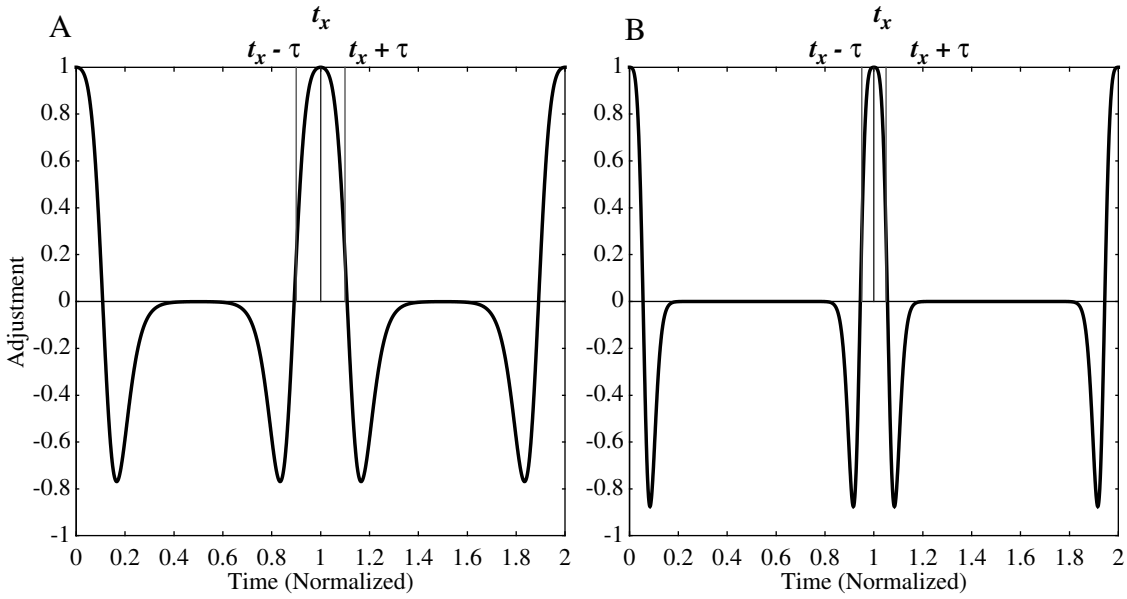


Figure 26: The effect of the delta rule for variability (τ) given in Equation 7 for two different values of τ. (A) $\tau = 0.10$. (B) $\tau = 0.05$. The y-axis gives $\Delta\Omega$ values, and τ is calculated from Ω according to Equation 6.

6.2.3.1 Confidence

Finally, it is useful for the oscillator to adjust the amplitude of its output pulses, providing an internal measure of performance. Confidence, c, can then be used to measure the success of the oscillator in finding pseudo-periodic event train in the input signal. There are a number of possible ways to do this. The most straightforward of these is to let confidence, c, be inversely related to variability, τ. Thus, as variability in the input signal shrinks, confidence grows. When confidence is calculated this way, no extra delta rule is needed; c can be limited to a fixed range between c_{min} and c_{max}, and modulated according to the value of Ω. One way to do this is given by the following equation, and this relationship is shown in Figure 27, for $c_{min} = -1$ and $c_{max} = 1$. Output may be defined to grow in amplitude with c.

$$c = c_{min} + 0.5\,(c_{max} - c_{min})\,(1 + \tanh\Omega)\,. \qquad \text{(Eqn 8)}$$

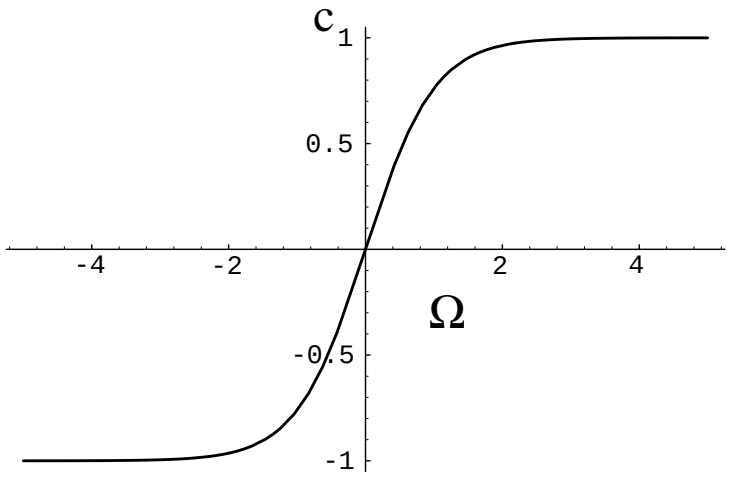


Figure 27: A possible relationship between Ω and c, according to Equation 8. This provides a measure of performance.

6.3 Oscillator Behavior

Figure 28 shows the behavior of an oscillator defined according the above equations, with initial conditions $p = 700ms$ and $t_x = 0$. The oscillator was exposed to input impulses with a period of $660ms$, shown in Panel (A). Model parameters were $\eta_1 = 1.0$, $\eta_2 = 0.3$, $\eta_3 = 0.3$, $\tau_{min} = 0.02$, $\tau_{max} = 0.50$, $c_{min} = -1$ and $c_{max} = 1$. The figure illustrates several aspects of the oscillator's behavior. Panel (B) shows the output pulses. In response to input, the oscillator adjusts its phase and period so that it becomes synchronized to the stimulus within a few cycles. As it synchronizes, output pulses shrink in width and grow in amplitude. Panel (C) gives more detail about this process. Observed cycle times of the oscillator are graphed with a solid line, capturing the combined effect of phase-tracking and period-tracking (Equations 4 and 5). Observed cycle times quickly adapt to the input cycle time of $660ms$. The dotted line in panel (C) shows intrinsic period, p, capturing the effect of the period tracking rule alone (Equation 5). The intrinsic period adapts more gradually than actual cycle time. Panel (D) shows the value of τ, the amount of variability the oscillator will tolerate in the phase and period of the input signal, as a percentage of the current period. When the stimulus is removed, the oscillator continues with a period of $660ms$. The oscillation at this new period may be said to embody an "expectation" for events at these particular future times.

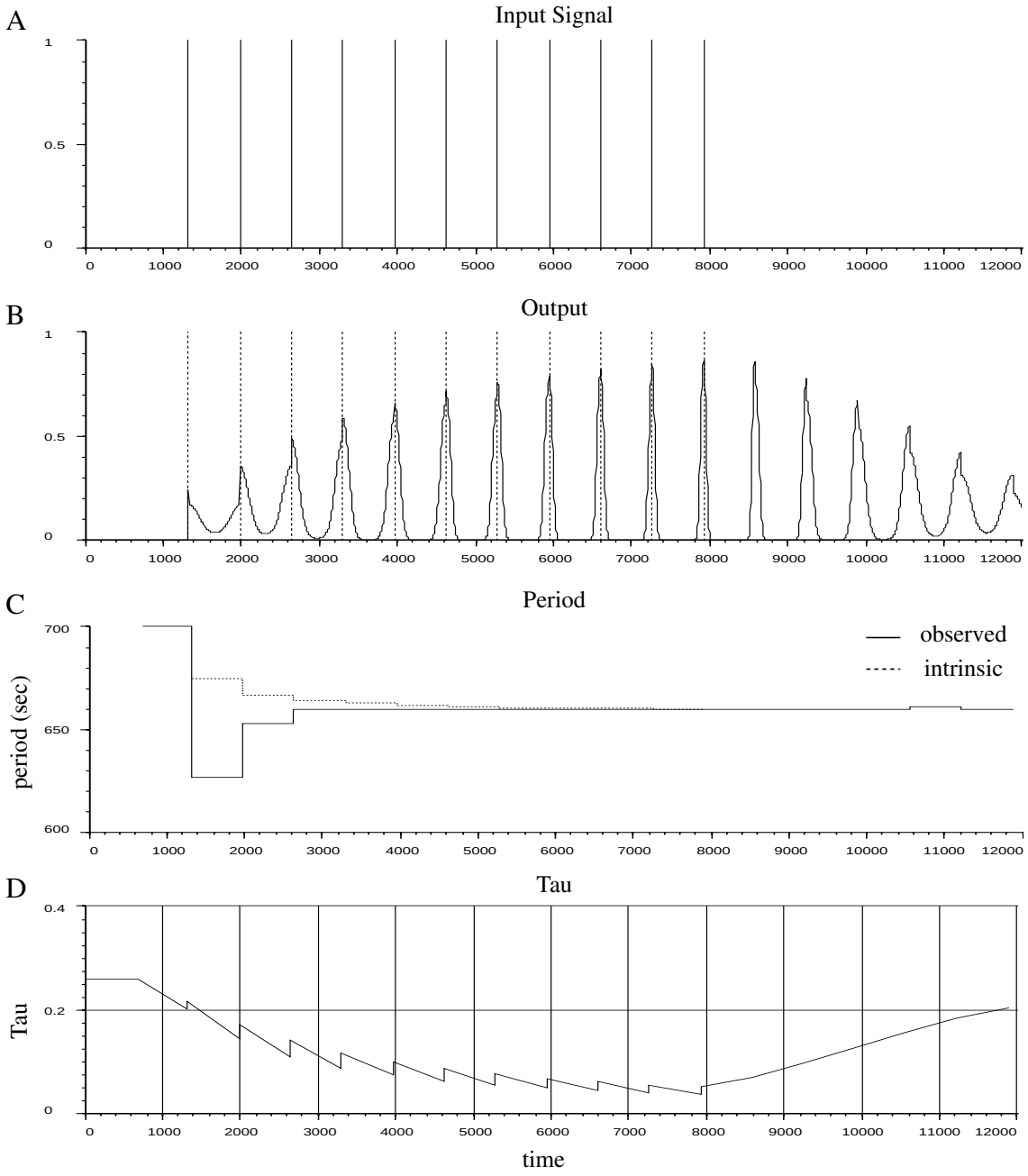


Figure 28: Single unit tracking a periodic signal: (A) input signal, (B) oscillator output pulses, (C) oscillator period, (D) τ.

6.3.1 Compound Units

Finally, in the examples that follow (Chapter VIII), each single unit is implemented using two oscillators operating in a tight, winner-take-all interaction. This arrangement is referred to as a *compound unit*. One oscillator, the *shadow* unit, is constrained to have the same period as the other, the *control* unit, but to remain exactly 180 degrees out of phase. The unit with the greatest confidence at any given time is defined to be the control unit. The control unit controls phase and period adjustments and produces output pulses. However, both units (shadow and control) actively adjust Ω, controlling τ, variability, and c, confidence. If at any time the shadow unit's confidence grows greater than that of the control unit, the two oscillators switch roles: the shadow becomes the control, and the control becomes the shadow. The observed behavior of a compound unit in this situation is a sudden 180 degree phase shift. This phase shift corresponds to a *gestalt* perceptual shift in the perception of the input signal – events that were perceived as out of phase with the beat are suddenly perceived as in phase. Figure 28 demonstrates a situation in which a *gestalt* perceptual shift may occur, $t = 3600ms$. Panel (D) in this figure graphs confidence of the control unit as a solid line, and confidence of the shadow unit as a dashed line. The gestalt perceptual shift happens when the two curves intersect. It also observable as a drop in cycle time, effective for a single cycle. Intrinsic period is not affected. The usefulness of the arrangement will become apparent in Chapter VIII, as specific examples are investigated.

6.4 Discussion

The oscillator synchronizes output pulses to a pseudo-periodic train of discrete impulses marking event onsets. Each output pulse instantiates a temporal receptive field for the oscillatory unit – a window of time during which the unit "expects" an impulse. The unit responds to impulses that occur within this field by adjusting its phase and period, and ignores stimulus pulses that occur outside this field. The oscillator adjusts the width of its receptive field to entrain efficiently. The oscillator entrains 1:1 to a simple periodic event train. A metrical structure, however, consists of levels of beats with different periods. The following chapters address this and related issues.

The approach provides a method for analyzing a complex rhythmic pattern as a set of pseudo-periodic components, or event trains. I have described the model as a single abstract oscillator, in order to focus attention on the adequacy of the proposal for modeling the human response to musical rhythm. This approach could be implemented in several ways. It is also possible to modify the delta rules slightly to handle markers that are not discrete impulses, but have shape and extent in time. It is important that the phase and period are tracked in such a way that only those events that fall within a temporal receptive field affect the behavior of the unit.

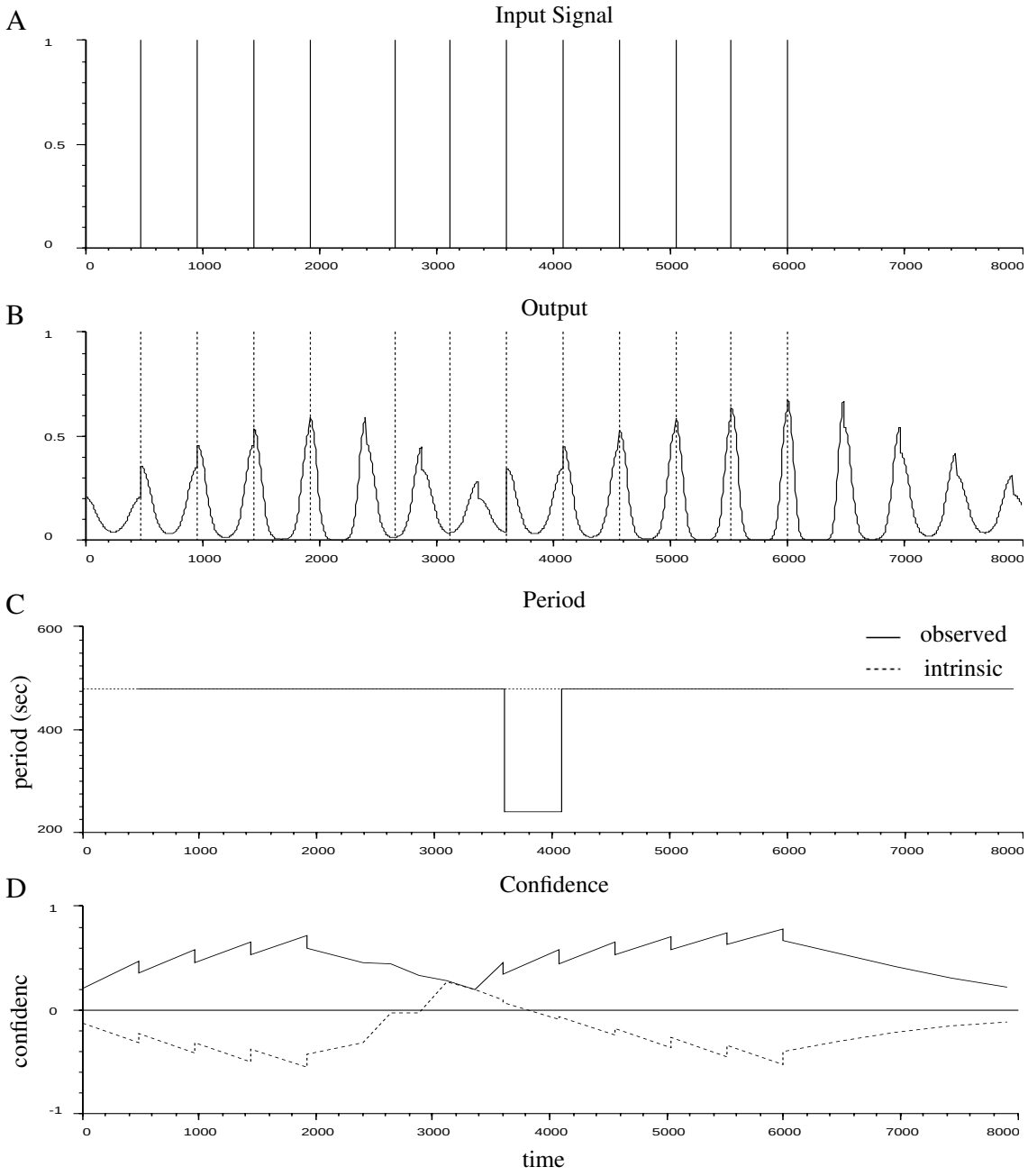


Figure 29: Single unit tracking a periodic signal: (A) input signal, (B) oscillator output pulses, (C) oscillator period, (D) τ.

CHAPTER VII

MODELING BEAT PERCEPTION AS A DYNAMICAL SYSTEM

The preceding chapter presented two dynamical systems, an event generator and an oscillator, and a way of coupling the two systems together, a set of delta rules. Examples of the behavior of the coupled system showed that the oscillator can entrain to a stationary signal whose period is close to its initial period. In the general case analysis of such a system may be quite complex, because the input signal may not be periodic, or its period may not be near the oscillator's initial period. The first step toward understanding the behavior of the coupled system is to make a geometric model of the states of the system. That is the goal of this chapter. This chapter will develop a dynamical system model of beat perception for the special case of an isochronous input signal. Development of this model will also have a useful side-effect: an efficient algorithm for simulating the behavior of a coupled system in the general case.

7.1 The Sine Circle Map

If two self-sustaining oscillators are physically separate, such that the behavior of one is not influenced by the behavior of the other, they are called *uncoupled* (Abraham & Shaw, 1992). The state space of each oscillator can be reduced to its limit cycle, a circle in the plane, and the state of each oscillator can be summarized by an angle identifying a position on the limit cycle. The combined state of the two oscillators may then be described as a pair of phases, θ and ϕ, yielding a single state space for the combined system. The state space is the cross-product of the two limit-cycles, topologically equivalent to the surface of a torus as illustrated in Figure 30.

The trajectory of a point in this state space, corresponding to the phase of each oscillator, winds around the torus. The two components of the trajectory correspond to θ, winding around the major axis (the "doughnut hole") and ϕ, winding around the minor axis (the "waist") of the torus (Abraham & Shaw, 1992). If the trajectory does not close on itself there is phase drift, and the trajectory will eventually cover the complete surface of the torus. This may be the case when two oscillators are completely independent. The resulting motion is called quasiperiodic (Abraham & Shaw, 1992). If the trajectory closes on itself after an integer number of cycles, then the two processes are *synchronized*. The resulting motion is periodic.

An interesting case is given when the behavior of one process (called the "driver", or the "driving oscillator") affects the other (called the "driven oscillator"), such as might be caused by a

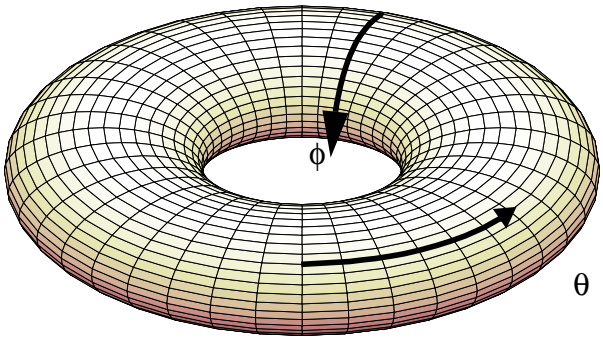


Figure 30: The state space for a system of two oscillators, a torus. A position on the surface of the torus describes a the combined system as a pair of phases.

slight mechanical connection between two oscillators. Now the two oscillators are *coupled*, and may become synchronized by a process called *entrainment*. Entrainment occurs because *coupling* allows the driver to perturb the driven oscillator by altering its phase, its intrinsic period, or both. The torus is also the state-space for the coupled system. Coupling means that the phase portrait is perturbed by the addition of small vectors at each point in this state space (Abraham & Shaw, 1992).

The effect of coupling may be understood by examining the trajectory of the combined system on the torus. To simplify the description one can slice the torus, taking a Poincare' section of the state space. This technique is analogous to observing the phase of the driver with a strobe that fires just as the driven oscillator passes $\phi = 0$, sampling the phase of the driver only at those times. The Poincare' section corresponds to a finite-difference equation, a one-dimensional discrete-time map from points on the circle to points on the circle. This map is called a Poincare' map, or simply a circle map. Such a difference equation describes the phase of the driver at which the driven oscillator will fire on its next cycle. The circle map provides a way to calculate the long term behavior of a system of coupled oscillators.

Consider the following mapping:

$$\theta_{i+1} = \theta_i + \frac{p}{q} + b \sin(2\pi\theta_i) . \qquad \text{(Eqn 9)}$$

This equation is a model circle map, called the *sine circle map*, that describes the dynamics of a system of two oscillators, a driving and a driven oscillator. The parameter q is the period of the driving oscillator, p is the period of the driven oscillator, and $b \sin(2\pi\theta_i)$ is a non-linear coupling term that describes the perturbations delivered to the period of the driven oscillator by coupling to the driver. θ_i is the phase of the driving oscillator at which the driven oscillator fires on iteration i. Figure 30 graphs this finite difference equation for $p/q = 2/3$, and $b = 0$. When $b = 0$ (no coupling), the behavior of the system is summarized by the ratio p/q, the so-called "bare winding

number". For example, if $p = 2$ and $q = 3$, the driven oscillator fires three time as the driver fires twice.

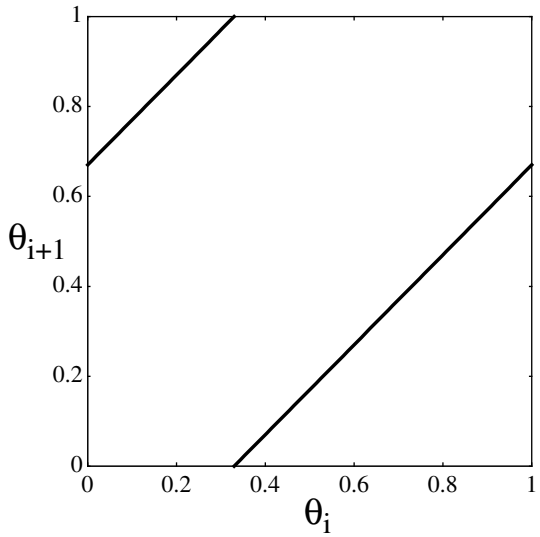


Figure 31: A graph of the finite difference equation given by Equation 9 for $p/q = 2/3$, and $b = 0$.

As coupling strength, b, increases, another ratio, N:M, the so-called "dressed winding number", describes the long-term behavior of the system. In the dressed winding number, N is the period of the driven oscillator under the influence of coupling, and M is the period of the driver. If the coupling strength is high enough, even as p/q is perturbed away from 2/3, the system will still lock in a 2:3 relationship, because each time the driven oscillator fires, its phase is perturbed slightly by the coupling to the driving oscillator (Glass & Mackey, 1988).

This locking behavior is highly structured. The dynamics of coupled systems like the sine circle map can be summarized in a *regime diagram*. Equation 32A shows a regime diagram for the sine circle map. The x-axis is the bare winding number, p/q, and the y-axis is coupling strength, b. The regime diagram identifies stable phase-locked states, also called attractors, mode locks, or resonances (Schroeder, 1991), for particular coupling strengths and driven/driver period ratios. The parameter regions that correspond to stable phase-locked states are known as Arnol'd tongues (Glass & Mackey, 1988; Schroeder, 1991). Each "tongue" is labeled with a ratio corresponding to its locking mode. The width of each tongue reflects the stability of the corresponding mode lock for a given coupling strength, i.e. its sensitivity to noise in the p/q ratio. Equation 32 shows that, for a fixed coupling strength, 1:1 entrainment is more stable than 1:2 entrainment, which is more stable than 2:3 entrainment, and so forth. Depending upon the coupling strength, it can be shown that entrainment is possible at any frequency ratio, N:M where N and M are relatively prime integers (Glass & Mackey, 1988).

The regime diagram is not arbitrarily organized. Its structure can be summarized by a mathematical construct known as the Farey tree (Equation 32B). The Farey tree enumerates all

rational ratios according to the stability of the corresponding mode lock in the coupled system. Its branching structure corresponds the structure of the Arnol'd tongues of the sine circle map, as well as to known bifurcation routes in other mathematical and natural systems (Schroeder, 1991). Regime diagrams and Farey tree have been used to model and predict biological and psychological phenomena (Glass & Mackey, 1988; Treffner, & Turvey, 1993; Schmidt, Shaw, & Turvey, 1993; Beek, Peper, & van Wieringen, 1992).

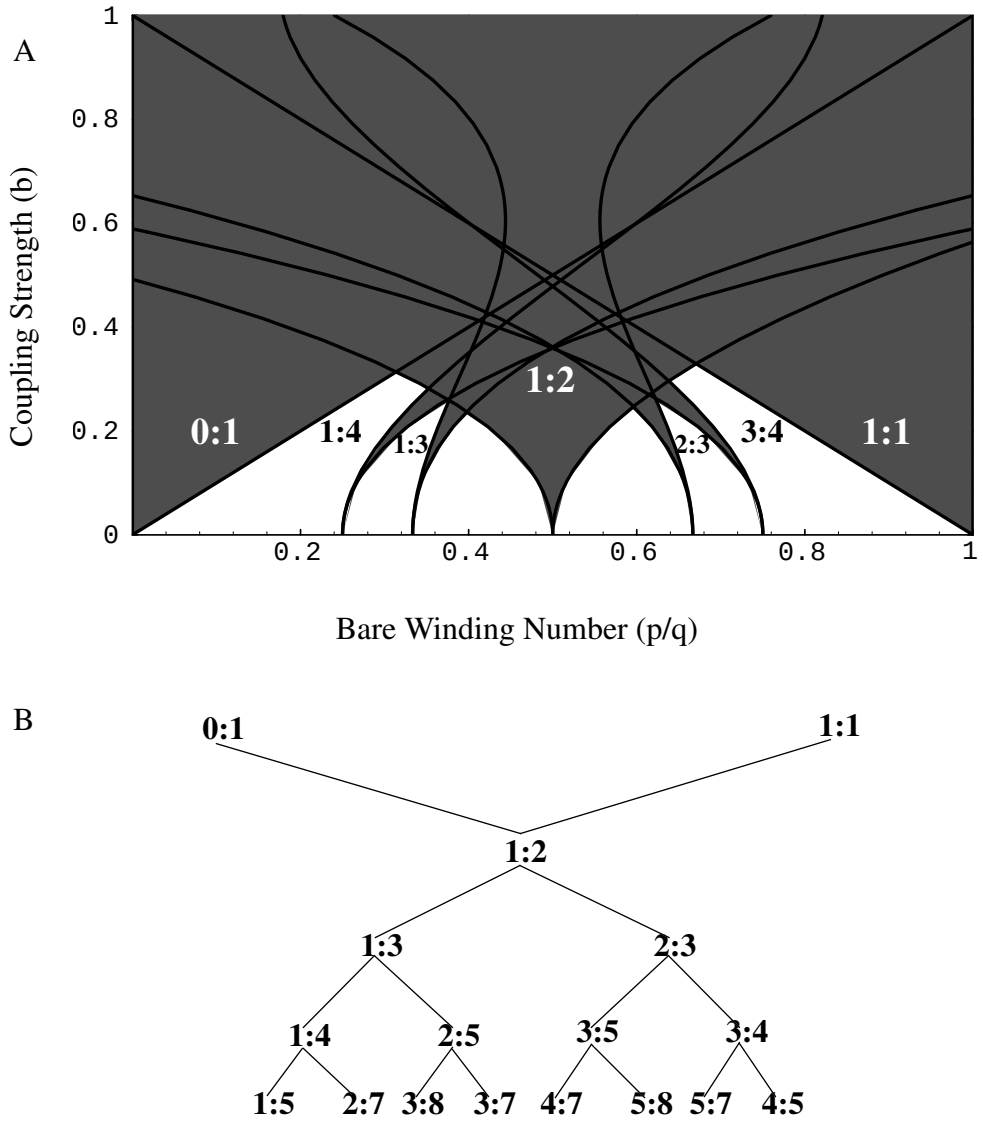


Figure 32: An Arnol'd tongues diagram (A) and the Farey tree (B).

7.2 Adapting the Circle Map

The beauty of the techniques described in the preceding section is that any two dynamical systems may be combined into a single system by taking the Cartesian product of their state spaces (Abraham & Shaw, 1992). The combined system can then be studied using the geometric techniques introduced above to arrive at a global understanding of the effect of coupling on observed behavior. The preceding chapter presented two dynamical systems, an event generator and an oscillator, and a way of coupling the two systems together, a set of delta rules. In the general case analysis of this combined system may be quite complex, because the event generator may not be a periodic process, so the simple toroidal state-space (as described above) would not be adequate. In this section, I assume that the event generator *is* periodic – an assumption that will provide a start at understanding the behavior of the oscillator model. A dynamical system on the torus will be derived as a geometric model of beat perception in this simplified case. I will use regime diagrams to study the behavior of the coupled system. This analysis will be useful in understanding how the unit will respond to any periodic driving stimulus, whether that stimulus arises from an external signal, or from the output of another oscillator in a network. Thus, the analysis will provide insight into several key aspects of the model. This simplified analysis will also have a useful side-effect: an efficient algorithm for simulating the behavior of the system, not only for the simplified case, but also in the general case.

7.2.1 Phase-Coupling

In the model of musical beat proposed in the previous chapter, the driver is a rhythmic pattern. Impulses in the signal perturb both the phase and the period of the driven oscillator through the action of delta rules. The delta rules provide a non-linear coupling that allow the signal impulses to affect the behavior of the oscillator. The oscillator adjusts its phase and period only at certain points its cycle, isolating individual event trains in the incoming signal.

Because of the simplifying assumption that the input signal is periodic, most of the assumptions about state spaces and trajectories from the previous section hold for this new coupled system. The heart of this analysis will be the formulation of a new circle map, a finite difference equation that summarizes the phase-tracking behavior of the oscillator in response to a periodic input signal. The basic form of the difference equation will be the same as that of the sine circle map,

$$\theta_{i+1} = \theta_i + \frac{p}{q} + f(\theta_j). \qquad \text{(Eqn 10)}$$

Iteration of this equation will allow study of the long-term behavior of the coupled system. The task is to determine the coupling term, $f(\theta_j)$ for this equation. The coupling term is derived from Equation 3 (pg. 86), the delta rule implementing phase-tracking.

In the sine circle map, θ_i is the phase of the driver at which the driven oscillator fires. However, for this new system a difficulty arises. If we strobe the driver when the driven oscillator

fires, we may see nothing at all, because the only times at which $s(t) > 0$ are when there is an impulse in the input signal. This is because Equation 4 ensures that the value of Δt_x will only be non-zero only when an impulse occurs in the signal. Therefore, the circle map must be adapted to suit current purposes. According to Equation 4, $\phi(t)$ is the phase of the driven oscillator at which an impulse occurs in the signal. This is the quantity upon which the amount of adjustment is based. So to develop a circle map describing the behavior of this model, the driven oscillator is strobed by impulses in the driving signal. Relative phase will be denoted by ϕ_j.

Assume a signal generator with period q, that generates a driving signal (a discrete series of impulses). Let T_j be the times at which the driver fires and let t_i be the times at which the driven oscillator fires, equivalent to the series of t_x's generated by the oscillator of the previous chapter. Let ϕ_j be the phase of the driven oscillator at which the driver fires. When $t_i \le T_j < t_{i+1}$, the phase of the driven oscillator at time T_j (the j^{th} signal impulse) is given by Equation 11. This situation is shown in Figure 33.

$$\phi_j = \frac{T_j - t_{i+1}}{p}. \qquad \text{(Eqn 11)}$$

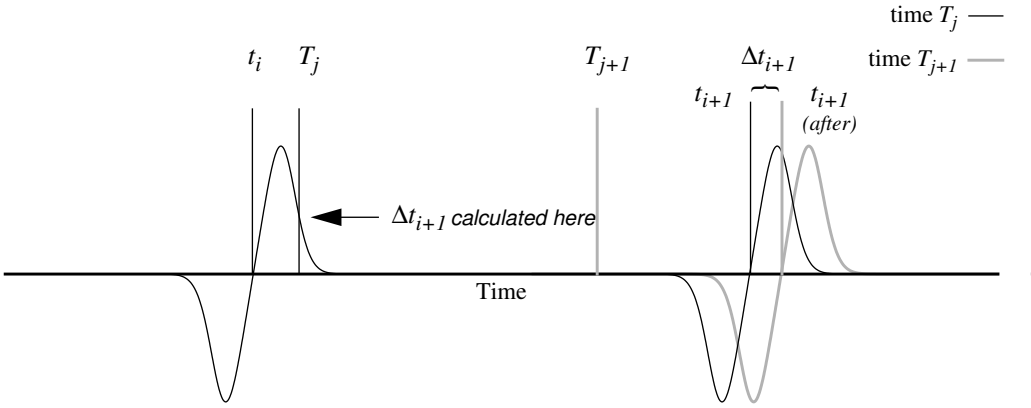


Figure 33: Stimulus impulse times, oscillator expected onset times and phase tracking delta rule. Solid lines show the situation up to time T_j, dotted lines show the situation after time T_j: the impulse causes a change in driven oscillator cycle length.

Equation 4 can be rewritten to give the change to the next expected time (Δt_{i+1}) after the onset of the j^{th} impulse:

$$\Delta t_{i+1} = \frac{\eta_1 p}{2\pi} \operatorname{sech}^2 \gamma (\cos 2\pi\phi_j - 1) \sin 2\pi\phi_j. \qquad \text{(Eqn 12)}$$

Now, substituting into Equation 1 relevant values of the variables ($t = T_{j+1}$, $t_x = t_{i+1}$, from Figure 33), the phase of the driven oscillator at which the next impulse will occur is:

$$\phi_{j+1} = \frac{T_{j+1} - (t_{i+1} + \Delta t_{i+1})}{p}. \qquad \text{(Eqn 13)}$$

Because the driver is periodic, $T_{j+1} = T_j + q$, and Equation 13 can be rewritten as:

$$\phi_{j+1} = \frac{T_j - t_{i+1} + q - \Delta t_{i+1}}{p}$$

$$= \phi_j + \frac{q}{p} - \frac{\Delta t_{i+1}}{p} \qquad \text{(Eqn 14)}$$

and this gives the necessary solution:

$$\phi_{j+1} = \phi_j + \frac{q}{p} - \frac{\eta_1}{2\pi} \operatorname{sech}^2(\gamma \cos 2\pi\phi_j - \gamma) \sin 2\pi\phi_j. \qquad \text{(Eqn 15)}$$

There is an important difference between this circle map and the sine circle map of Section 7.1. Equation 15 is really an approximation. One cannot calculate the actual phases at which the driver fires between *i* and *i+1* until after the *i+1*st firing of the driven oscillator, because in fact *p* may change every time the driver fires – the firing of the driver alters the phase, and thus the period on that cycle. However, this circle map is appropriate for two reasons. First, it is an accurate portrayal of the operation of the system, because this approximate phase is the measurement of phase upon which the delta rule for phase-tracking is based. Second, as the coupled system converges on a stable limit cycle, the error in Equation 15 approaches 0.

There is an also important similarity between this circle map and the sine circle map. Equation 15 has an important special case when $\gamma = 0$. In this case, $\operatorname{sech}^2(\gamma \cos 2\pi\phi_j - \gamma) = 1$ and Equation 15 is the same as Equation 9, with the roles of driver and driven reversed. In a sense Equation 15 can be thought of as an extension of the sine circle map. An examination of the Arnol'd tongue diagram corresponding Equation 15 for values of $\gamma > 0$ (Figure 34, below) reveals the nature of this extension.

Rather than solving Equation 15 to analytically determine the boundaries of mode-locked states (as in Figure 32A), this difference equation is repeatedly iterated for different initial values of q/p and η_1, looking for limit cycles. This allows calculation of the number of cycles required for the system to converge. This information is useful because in actual cases the behavior of interest corresponds to the dynamical system's transients, not to its limit behavior. Thus, this information helps in understanding real-time performance. For each of the following regime diagrams, it is assumed that the driver and the driven oscillator initially fire together, so $\phi_0 = 0$.

Iteration of Equation 15 yields the regime diagrams of Figures 34, 35, and 36. These figures show stable resonance modes for rational ratios, q/p such that $p \leq 8$. Darker regions correspond to regions of faster convergence. Each individual picture corresponds to a different value of γ. Figure 34, the regime diagram for the coupled system with $\gamma = 0$, shows the relationship between this circle map and the sine circle map (compare Figure 34 with Figure 32). Figures 35 and 36 show entrainment zones for $\tau = 0.10$ and $\tau = 0.05$, respectively. As the diagrams show, the effect of

shrinking τ (increasing γ), thereby shrinking the oscillator's temporal receptive field. Zones of 0:1 and 1:1 entrainment shrink, allowing widening of the regions corresponding to more complex ratios. This allows the oscillator to acquire stable phase-locks in complex ratios with the input signal more easily.

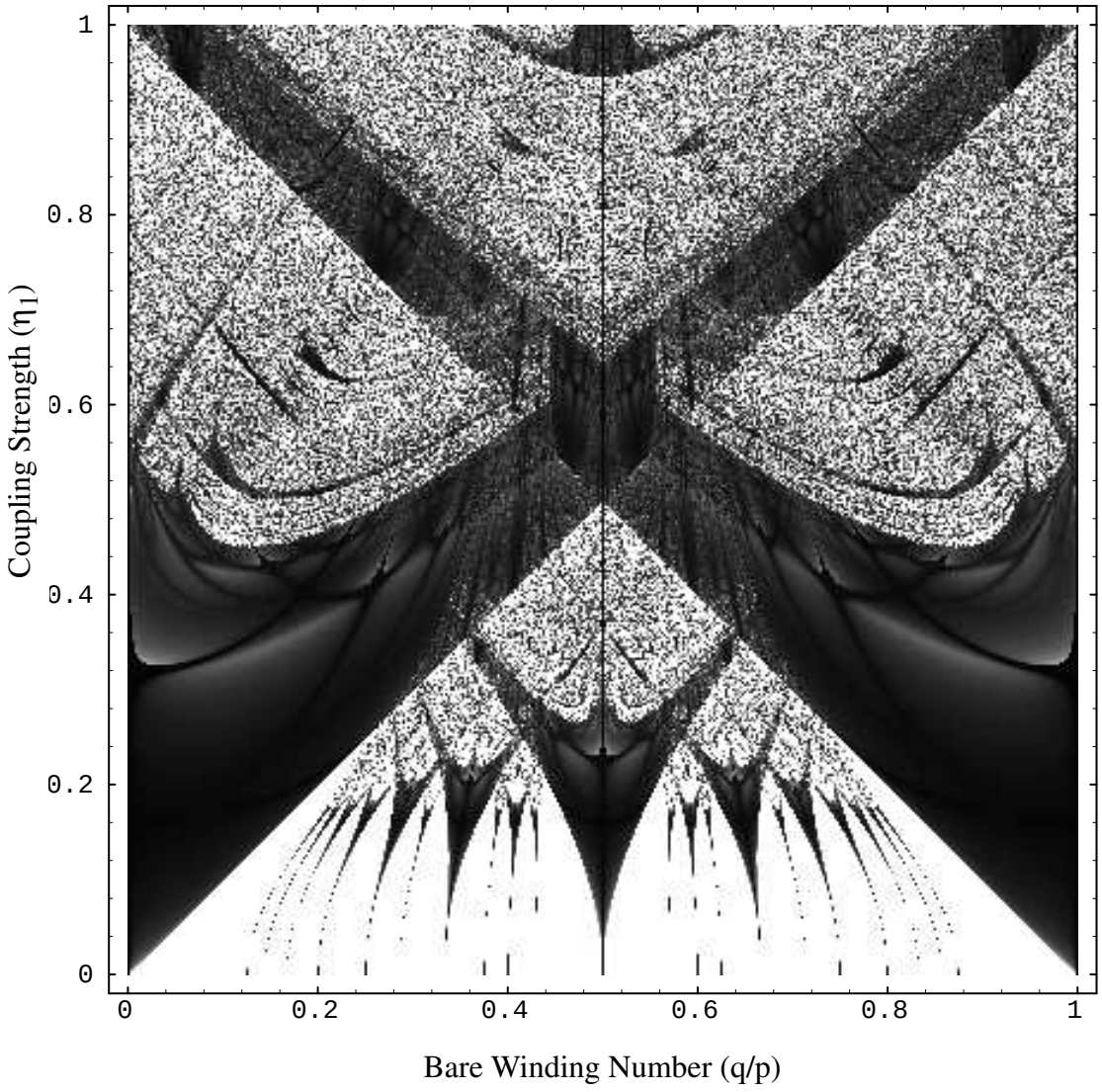


Figure 34: An empirical regime diagram for the phase-coupled model with $\gamma = 0$.

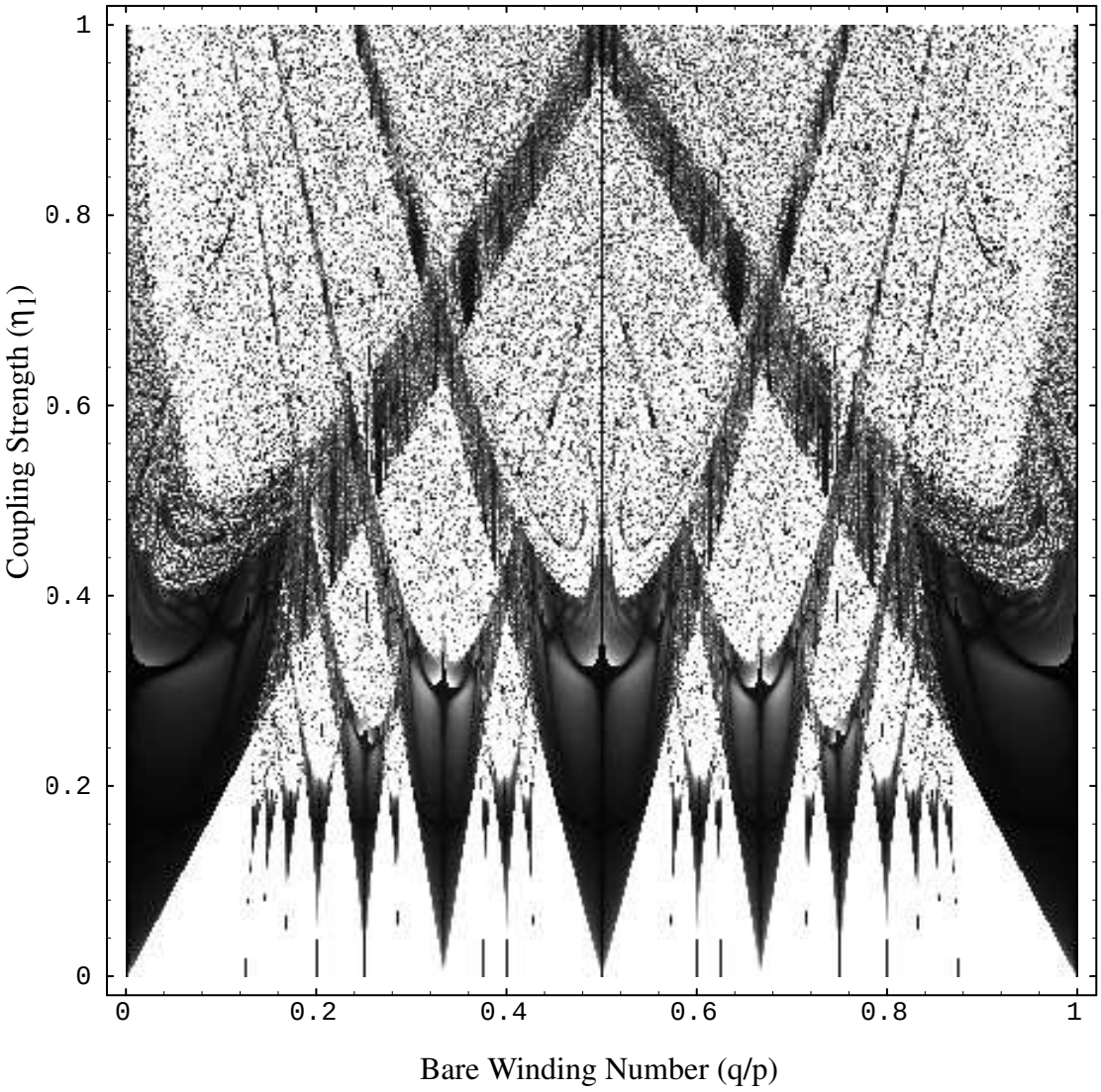


Figure 35: An empirical regime diagram for the phase-coupled model with $\tau = 0.10$.

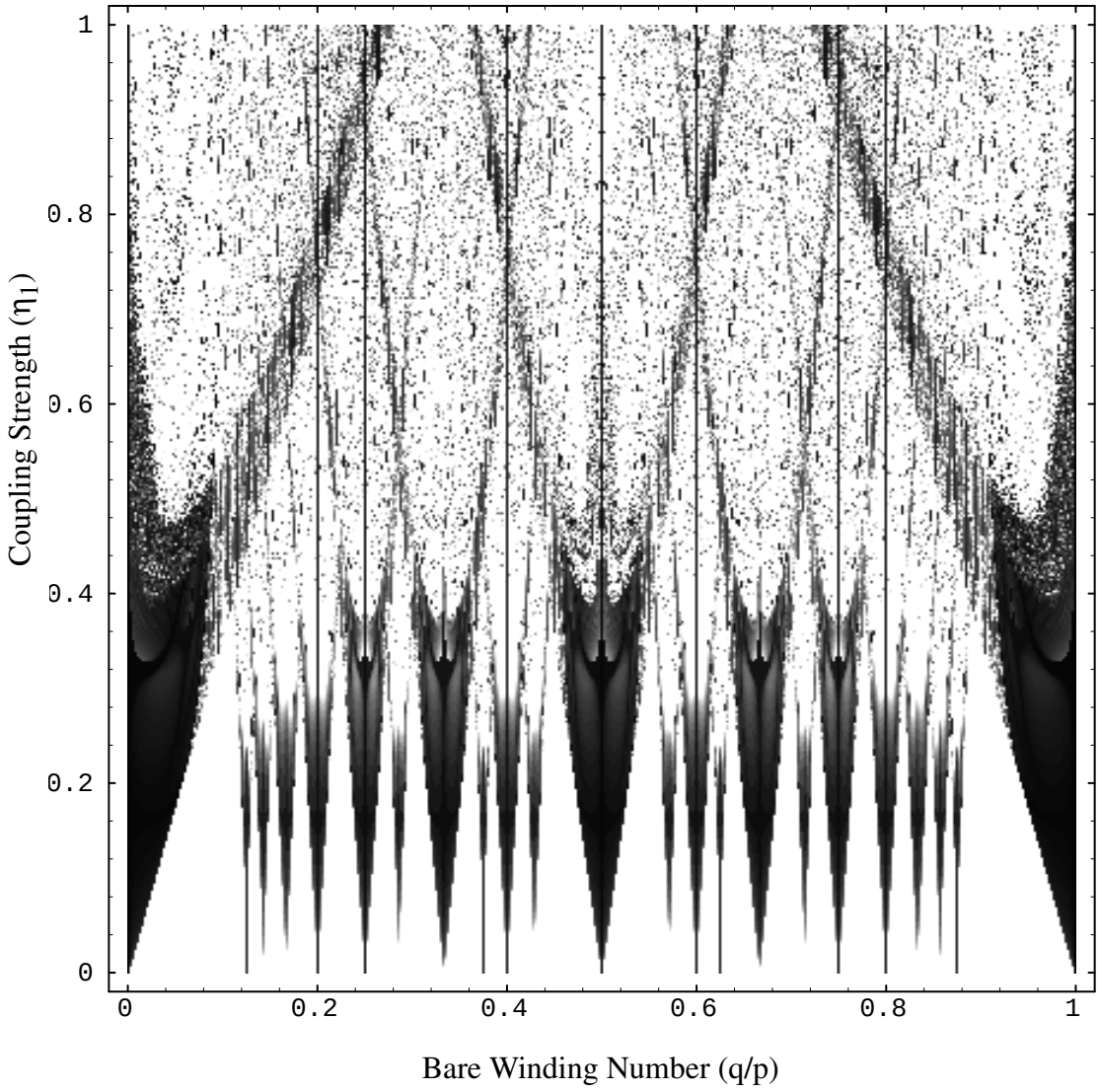


Figure 36: An empirical regime diagram for the phase-coupled model with $\tau = 0.05$.

7.2.2 Period-Coupling

In phase-coupled systems, the period of the driven oscillator is altered because its phase is perturbed in every cycle. When the effect of the driving signal is removed, even for one cycle, the driven oscillator reverts to its intrinsic period. When the driver returns, a number of cycles may be required to reestablish phase lock. As discussed in previous chapters, this behavior is insufficient for modeling musical beat: the oscillator model must also identify and remember the beat period. The oscillator of Chapter VI does this with a period-tracking delta rule that allows the driving signal to perturb the intrinsic period of the driven oscillator. The period-tracking oscillator can model musical beat because when the driving signal is removed, the oscillator continues at the driver's frequency, "expecting" the driver's eventual return. The dynamical system for modeling beat perception is not simply phase-coupled, it is also period-coupled.

Regime diagrams for the period-coupled system can also be developed by adding the following equation to the model:

$$p_{j+1} = p_j + \eta_2 \operatorname{sech}^2(\gamma \cos 2\pi\phi_j - \gamma) \sin 2\pi\phi_j. \qquad \text{(Eqn 16)}$$

Equation 16 is derived from Equation 5, the period-coupling delta rule. Figures 37, 38, and 39 show resonance tongues for the phase- and period-coupled system for γ=0, $\tau = 0.10$, and $\tau = 0.05$, respectively. For easy comparison with Figures 34, 35, and 36, Equation 16 was added to the model with η_2 fixed at a value of 0.02, again varying η_1 (from Equation 15) along the y-axis. The entrainment regions for the phase- and period-coupled system are larger than the corresponding regions for the simpler phase-coupled system. Period-tracking causes a widening of the resonance tongues. Therefore, not only does period-tracking act as a sort of memory, as described in the Chapter VI, but it also enhances the stability of the oscillator's response in the presence of timing deviations.

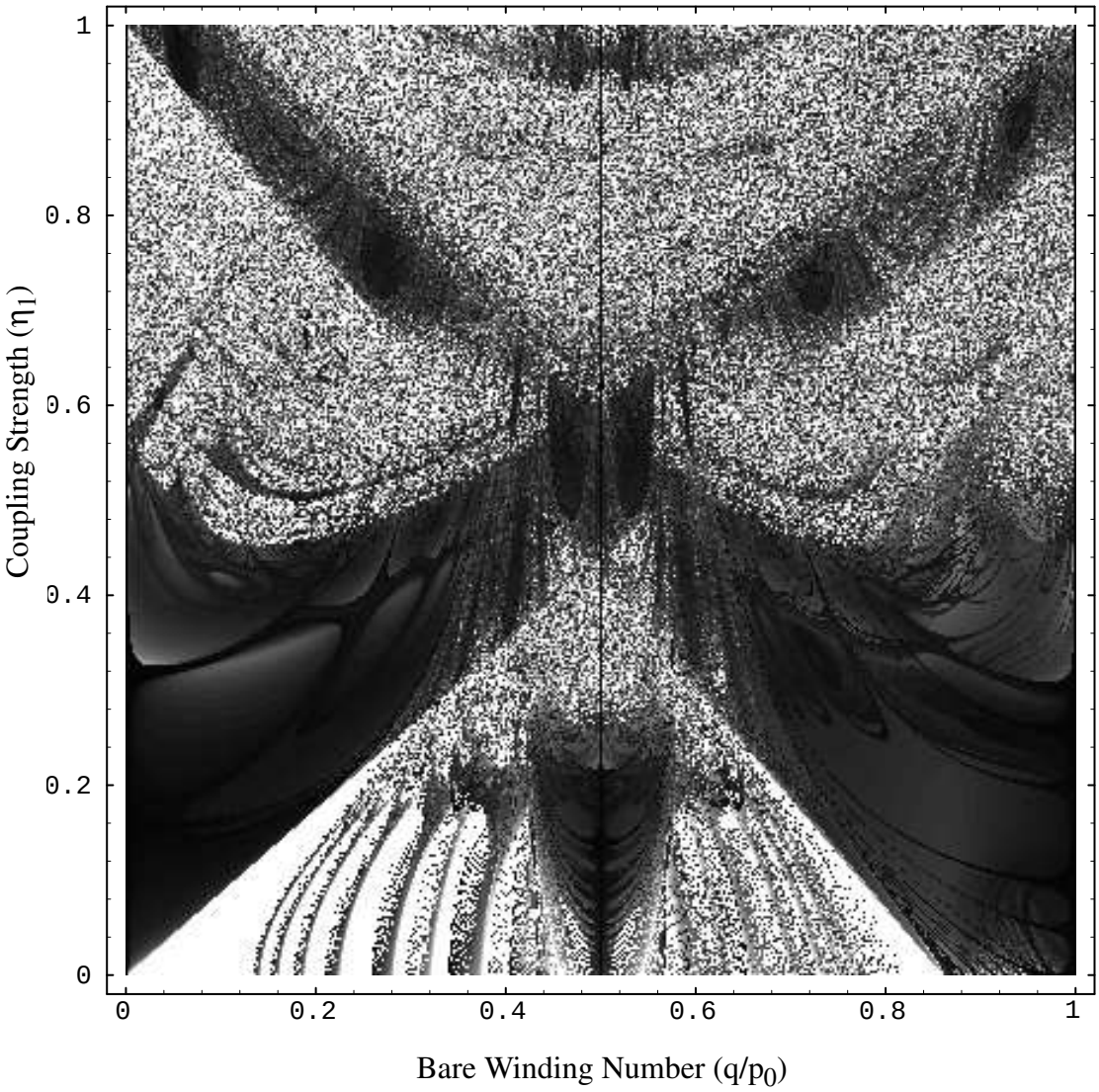


Figure 37: An empirical regime diagram for the period-coupled model with $\gamma = 0$.

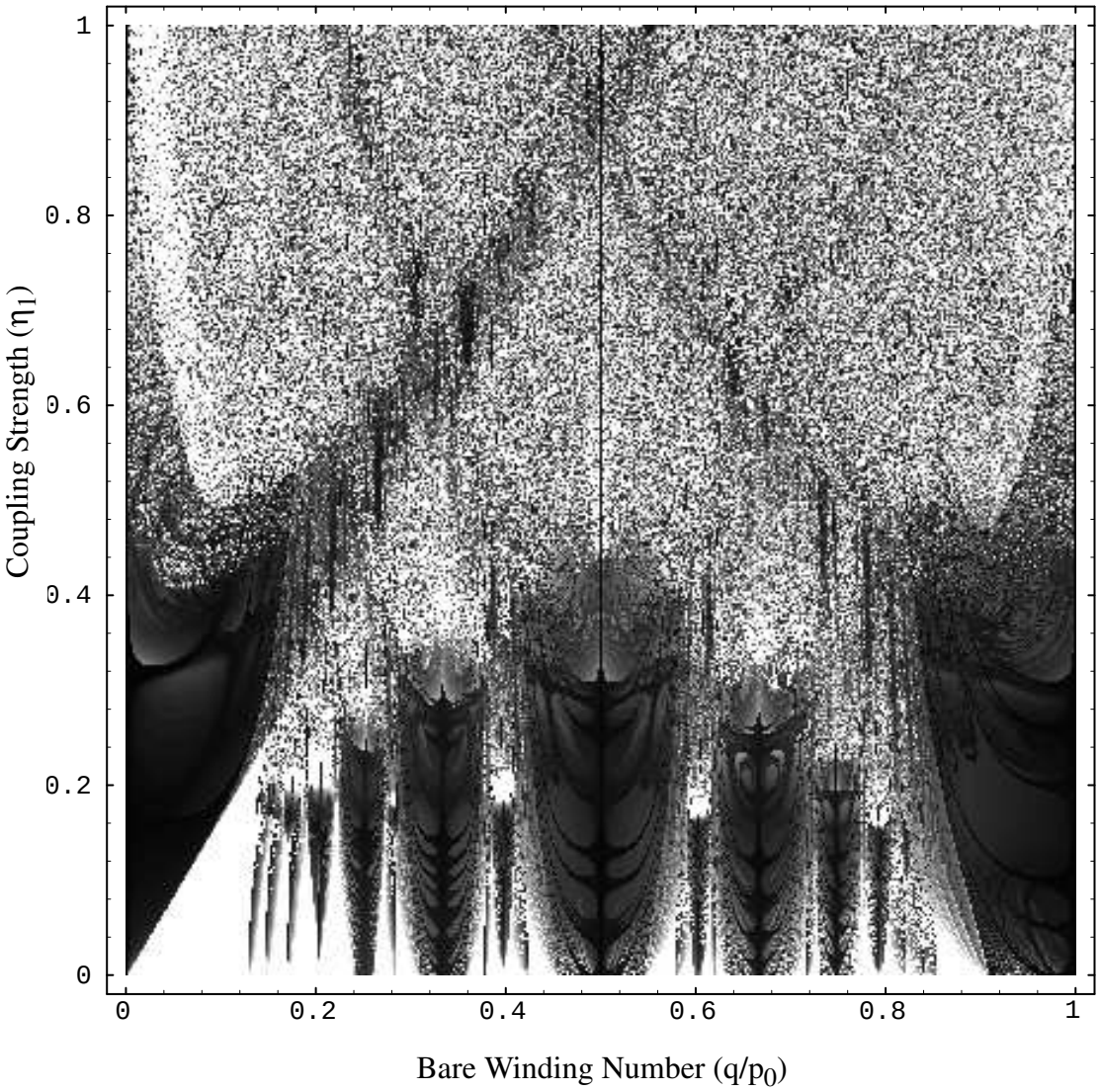


Figure 38: An empirical regime diagram for the period-coupled model, with $\tau = 0.10$.

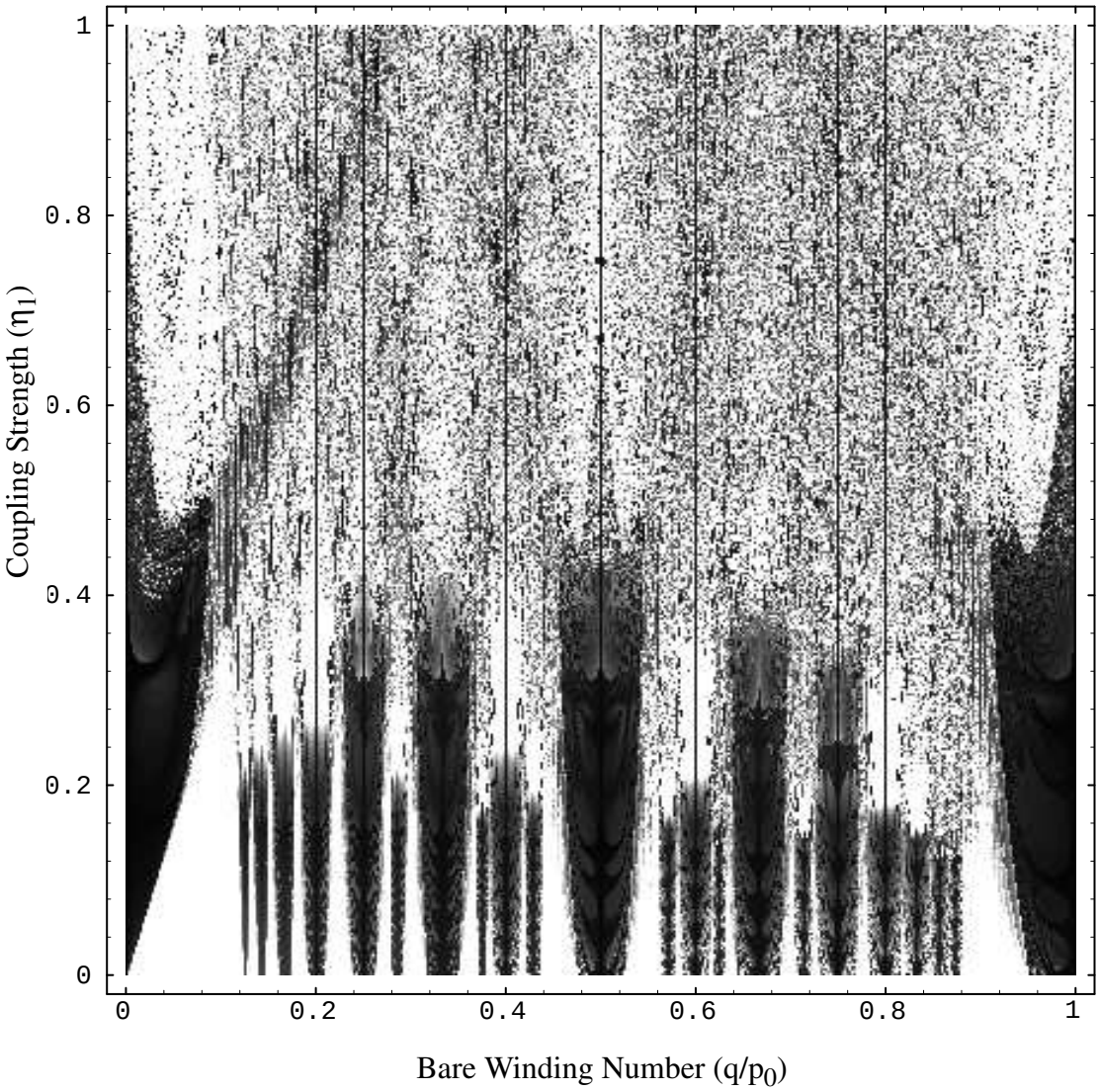


Figure 39: An empirical regime diagram for the period-coupled model with $\tau = 0.05$.

7.3 An Efficient Algorithm

To create a state-space for studying the coupled system, it was necessary to assume that driver was periodic. This was so that the state of the driver could be reduced to a circle and a toroidal state-space constructed for the coupled system. However, the difference equations derived for the circle map are actually quite general; only in actually calculating ϕ_j was it assumed that the driving signal was periodic. The driven oscillator was strobed whenever there was an impulse in the input signal. Therefore, as a side benefit of deriving the finite difference equations, an algorithm for calculating the oscillator's behavior has been created. Assuming that signal impulses and changes to the unit's parameters are discrete, this output of this algorithm is a time series that captures the unit's behavior in response to any signal, no matter how complex.

for t **from** 0 **to** end-of-signal (1)
  if $(s(t) > 0)$ **then** (2)
    $\phi = \frac{t - t_x}{p}$ (3)
    $t_x \leftarrow t_x + \eta_1 \frac{p}{2\pi} \operatorname{sech}^2 \gamma (\cos 2\pi\phi - 1) \sin 2\pi\phi$ (4)
    $p \leftarrow p + \eta_2 \frac{p}{2\pi} \operatorname{sech}^2 \gamma (\cos 2\pi\phi - 1) \sin 2\pi\phi$ (5)
    $\Omega \leftarrow \Omega + \eta_3 \operatorname{sech}^2 \gamma (\cos 2\pi\phi - 1)\ (\cos 2\pi\phi + 2\gamma \tanh \gamma (\cos 2\pi\phi - 1) \sin^2 2\pi\phi)$ (6)
    $\tau \leftarrow \tau_{max} + 0.5 (\tau_{min} - \tau_{max})\ (1 + \tanh \Omega)$ (7)
    $\gamma \leftarrow \frac{\omega}{\cos 2\pi\tau - 1}$ (8)
  endif; (9)
  if $(t = t_x)$ **then** (10)
    $t_x \leftarrow t_x + p$ (11)
    decay Ω (12)
    generate a beat; (13)
  endif; (14)
endfor; (15)

In this algorithm, line 4 is really just a version of Equation 12, and line 5 is really just Equation 15. Lines 6 - 8 implement a adjustment to γ by means of Ω and τ. The factor $s(t)$ can be added to the one or more of the delta rules if it is used to carry amplitude information. The algorithm as presented here, however, assumes that $s(t) = 1$ for every impulse, so its presence in the formula is redundant. This algorithm outputs discrete signals, or beats, in lines 10 - 14. The floating point operations inside the first conditional are executed only when a signal impulse is

present, and inside the second conditional when the oscillator fires. Hence this is an efficient algorithm, realizable in software in real time. Running time is linear in the number of signal impulses, with constant determined by the period of the oscillator.

7.4 Discussion

In this chapter a dynamical system model of beat perception was developed. To create the model of the coupled system, it was assumed that the input signal consisted of isochronous impulses. In fact, this assumption can be relaxed somewhat. For example, a finite-length input signal could be looped, and loop time equated with cycle time. This would be similar to assumptions made in computing Fourier transforms for finite length signals (Oppeheim & Schafer, 1975). The driven oscillator can be strobed by signal impulses to develop a circle map, because impulse times and phases are known in advance. Although such circle maps and their associated regime diagrams can be interesting and useful, the next chapter will take a different approach.

The most interesting behavior of this dynamical system is not its limit behavior, but its transient behavior. This is because music is not usually composed of precisely the same rhythm looped indefinitely; it is composed of rhythms that change. Development of this model yielded an efficient algorithm for simulating the behavior of the coupled system in this general case. To study the transient behavior of this coupled system, it is not necessary to calculate relative phase, but to simply calculate future firing times of the driven oscillator. The next chapter will use this method to study the transient behavior of dynmical systems that are created by coupling one or more oscillators to complex, temporally structured input signals.

CHAPTER VIII

SOME EXPERIMENTS WITH THE OSCILLATOR MODEL

Chapter VII studied the dynamical system that was created by coupling the oscillator of Chapter VI to an isochronous input signal. The behavior of the oscillator under the influence of coupling was interesting. The regime diagrams of Figures 34 through 39 showed complex responses to isochronous stimulation at different frequencies. Those diagrams, however, describe behavior that is simple by comparison with oscillators driven by complex musical rhythms. In this chapter, rather than creating a regime diagram for each dynamical system created by coupling an oscillator to a musical rhythm, transient responses are evaluated by examining times series corresponding to the response of individual oscillators to complex rhythmic patterns. Although regime diagrams make predictions about the limit behavior of such systems, the behavior most useful for the processing of temporal sequences is found in transient responses. These will allow study of how the oscillators respond to music-like signals in real-time.

This chapter describes two experiments. In the first experiment, systems are created by driving an oscillator with musical rhythms collected in the study of improvisational performance, reported in Chapter III. Melodies pose a good test for entrainment models, because they usually contain more rhythmic complexity than multi-voiced music. Bass lines and harmonic accompaniment, for example, are often metrically regular, providing more regular cues for entrainment than melodies. The driving signals are derived from performances, and contain systematic timing deviations. The goal of this experiment was to determine how well an oscillator can track an individual event train in an input signal, coping simultaneously with rhythmic complexity and systematic timing deviation. First performances of notated melodies and then performances of improvised variations are studied.

In the second experiment, stationary signals of varying levels of structural complexity (nested rhythms and polyrhythms) were used to drive a system of oscillators. The goal of this experiment was to determine whether it is possible to use a group of oscillators to identify metric relationships in music-like stimuli. Rhythmic complexity and temporal deviation were controlled for and the ability to identify metric relationships was studied in isolation. Both hierarchically nested structures (2:1 and 3:1) and polyrhythmic structures (3:2 and 4:3) were considered.

8.1 Performed Musical Rhythms

In this experiment, performances of notated melodies and performances of improvised variations were studied separately, because they differed qualitatively in level of rhythmic complexity, and

they differed significantly in the magnitude of timing deviations. Performances of notated melodies provided a controlled level of rhythmic complexity. Each melody contained three intended duration categories: sixteenth note, eighth note, and quarter note. This makes it easy to see how the oscillator deals with distractor events and missing event onsets. Improvised variations were rhythmically quite complex, making use of syncopation, and up to seven levels of intended duration categories (according to the transcriptions). These performances contain more distractors and missing events, providing difficult test cases for entrainment. Both types of performance contained timing deviations, making the task of tracking a single event train a challenging one.

A single compound unit (described in Section 6.3.1) attempted to track a single pseudo-periodic event train in the performances. The response of the oscillator was intended to model the perception of beats at some level in a metrical structure grid. For each performance, the modal inter-onset interval (IOI) category was determined from the score or transcription, and chosen as the target event train. The unit was initialized such that $\phi = 0$ at the start of the performance, and p was equal to the initial IOI of the target event train for that performance. Thus, the oscillator did not have to cope with finding initial phase or period.

Figure 40 gives an example of the oscillator's behavior as it tracks a performance of *Baa baa black sheep*. The first panel (A) provides a notated version of the melody (transcriptions of improvisations do not include grace notes or other ornaments) and a single row of dots from a metrical structure grid marking the target event train. Notes that are not marked by dots are to be ignored by the oscillator; dots that do not correspond to notes mark times when events are "missing" from the target event train. Panel (B) shows both input and output of the oscillator. The dashed lines show impulses in the input signal (marking event onset times), and solid lines show when the oscillator outputs beats ($t = t_x$). These two lines overlap when a target event is performed at precisely the time predicted by the oscillator, that is, at phase zero, $\phi(t) = 0$, of the driven oscillator. Amplitude of the oscillator output is controlled by confidence, c, providing a way for the oscillator to measure its own performance (Section 6.2.3.1).

Panels (C) and (D) both show the tempo curve for the performance as a solid line. This curve was derived by extracting the target event train from the performance and graphing IOI's for the target events. This curve gives the IOI's to which the oscillator should respond. Panel (C) also shows the oscillator's intrinsic period, adjusted by the period-tracking delta rule throughout the performance as a dotted line. For this performance, beginning at the initial tempo, the unit effectively calculates a local average tempo, following performance tempo as the performer speeds up and slows down, based on eighth note onset times only. Panel (D) shows actual observed cycle times of the oscillator using a dotted line. Observed cycle time takes into account not only the intrinsic period of the oscillator, but also the phase as it is adjusted by the phase-tracking delta rule in each cycle. Thus, this curve represents the combined effect of the two delta rules given by Equation 4 and Equation 5. Cycle time tracks the performance times much more closely than period alone. The last panel (D) shows τ, the size of the oscillator's temporal receptive field, as a percentage of oscillator period. At its lowest point in this performance, the value of τ is about 0.05. The curve is jagged, because each time the oscillator fires, the value of Ω (Equation 7) decays toward zero. The value of τ and oscillator confidence (the amplitude of the output beats) are inversely related. τ determines which onsets the oscillator will ignore.

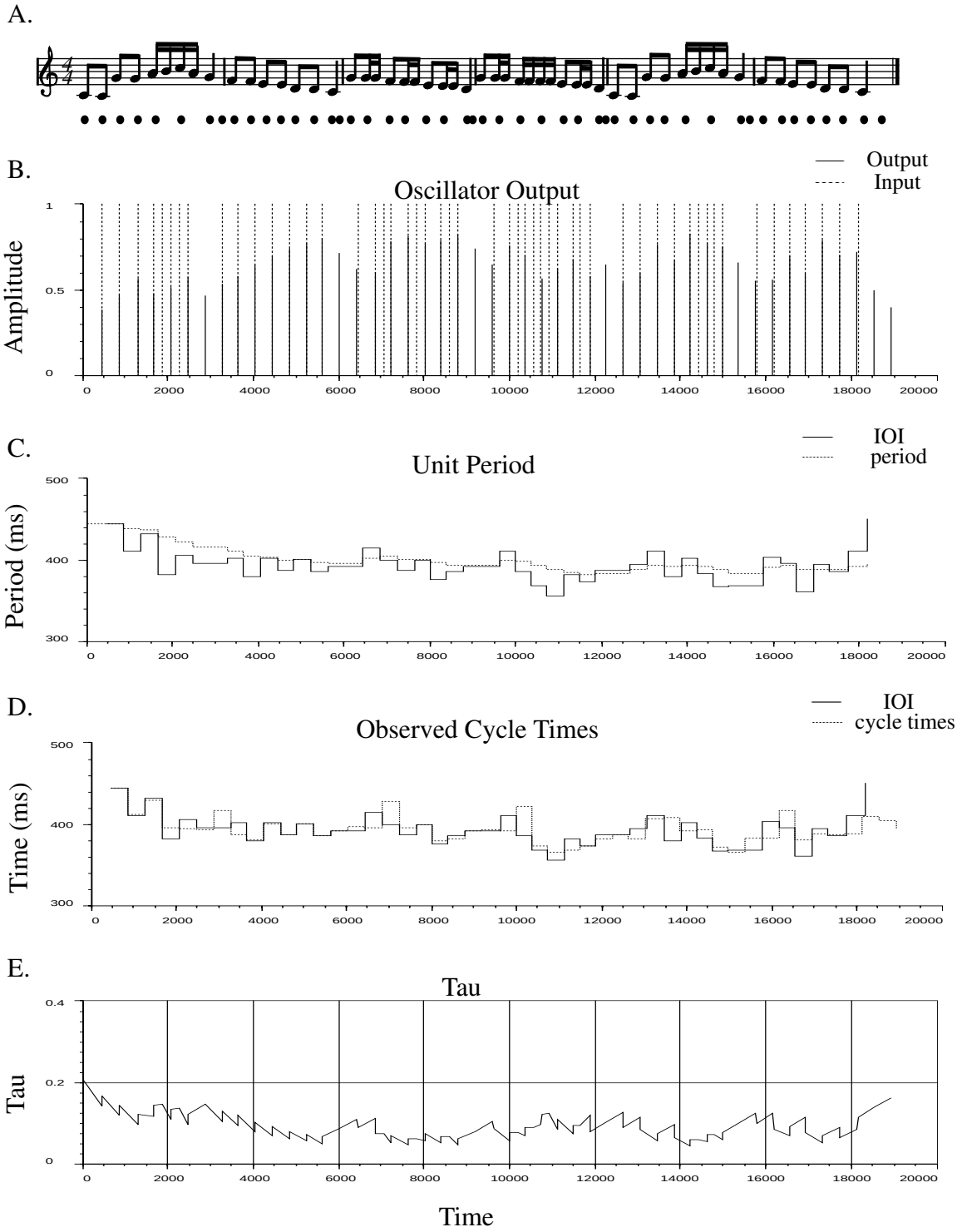


Figure 40: An oscillator tracking the rhythm of *Baa baa black sheep* (rubato = 0.05, $\overline{|\phi'|} = 0.08$; $R^2 = 0.34$, $p < 0.05$).

Performance at tracking events is measured as accuracy in predicting target onset times throughout a performance. Rather than examining the time series for each of the 60 performances, the oscillator's overall performance can be examined by computing summary statistics. The statistics will provide a measure of how well the oscillator performs on average, and they will help identify situations in which oscillator performance breaks down. Phase of the driven oscillator at which target events occur is an appropriate measure on this task. This measure is consistent with the dynamical system model of beat perception defined in the Chapter VII, that is, events in the signal act as strobes of the driven oscillator. Also, this number provides a measure of performance that can be compared with the measure of rubato collected in the initial study of Chapter III.

In Chapter III, rubato was defined for each performed inter-onset interval as the deviation from the average IOI for each IOI category in each performance. Comparison of oscillator performance with the rubato measure can be thought of as comparing the oscillator's performance on the prediction task with a simple default strategy: Given a target event occurring at time t, predict that the next target event will occur at time $t + m$, where m is the mean IOI for the target event train in the current performance. This is not a realistic strategy for predicting target event onset times, but it provides a baseline measure of how well the oscillator might be expected to perform, given the amount of timing deviation present in the performance.

To compare the measure of rubato with phase, the notion of phase must be modified to create a measure of deviation from expected target event onset time. In the dynamical system model, phase, ϕ, was defined to vary from 0 to 1. For this study, a new measure of phase, ϕ', is defined that varies from -0.5 to 0.5. $\phi' < 0$ when an event occurs earlier than expected, $\phi' > 0$ when an event occurs later than expected, and $|\phi'|$ can be used as a measure of deviation from expected target event onset time (as a proportion of oscillator period). For example, if the oscillator predicts a target event onset precisely, $|\phi'| = 0$, whereas if a target event occurs 180 degrees out of phase with the oscillator's prediction (halfway through the oscillator's period), $|\phi'| = 0.5$. This measure can be averaged over each performance and compared with deviation from mean tempo, as a measure of the oscillator's performance on the tracking task.

To assess the performance of the oscillator more carefully, a measure that evaluates unit performance on an event-to-event basis (in real time) is also necessary. Observed oscillator cycle times can be compared with actual performed IOI's, shown together in panel (D). Correlations between cycle times and target IOI's measure how closely cycle times track actual IOI's in real time. Significant correlations show that the oscillator is tracking near-optimally, while nonsignificant correlations indicate some difficulty.

Figure 40 showed an example of the oscillator tracking a notated performance well (rubato = 0.05, $\overline{|\phi'|} = 0.08$; $R^2 = 0.34$, $p < 0.05$). In the following analyses, summary statistics will be used to tally cases in which the oscillator's performance is good (i.e. roughly comparable to performance in Figure 40). Difficult cases are then identified and examined in detail to learn how performance breaks down, and under what circumstances. First performances of notated melodies and then performances of improvised variations are considered.

8.1.1 Performance of Notated Melodies

In each of the notated melodies, there were three IOI categories, corresponding to sixteenth note, eighth note, and quarter note durations. The modal duration category (eighth notes in each melody) was chosen as the target to test oscillator performance. Sixteenth note IOI's corresponded to distractor events, and quarter note IOI's corresponded to "missing" onsets in the target event train.

The unit was exposed to thirty melodies collected in the initial study – five performances each of three melodies by the two pianists included in the timing analyses of Chapter III. $|\phi'|$ was averaged over each performance, yielding $\overline{|\phi'|}$ as a measure of the oscillator's performance in the tracking task. An analysis of variance (ANOVA) was conducted with factors melody, pianist, and analysis type (rubato vs. phase). The ANOVA showed a main effect of analysis type ($F(1, 4) = 27.73$, $p < 0.01$), with mean rubato = 0.05, and average phase = 0.06. This shows that for these performances, the oscillator did not perform as well as the baseline strategy of predicting the next target event onset time based on mean IOI for the target event train. This value of $\overline{|\phi'|}$, however, indicates that on average the oscillator is tracking the target event trains well. The ANOVA also showed a significant main effect of subject ($F(1, 4) = 11.19$, $p < 0.05$). To assess the performance of the oscillator more carefully, observed oscillator cycle times were compared with actual performance times. Correlations between oscillator cycle times and performed IOI's within each performance were significant ($p < 0.05$) for 22 out of the 30 melodies. The eight cases in which the correlations were not significant are cases in which the unit may be having difficulty tracking the signal. These difficulties are examined in detail by investigating two representative cases.

8.1.1.1 Case 1

Figure 41 is one such case, a performance of *Hush little baby* (rubato = 0.07, $\overline{|\phi'|} = 0.04$; $R^2 = 0.32$, $p = 0.29$). In this case, there is a disparity between statistical measures of performance. The low value of $\overline{|\phi'|}$ shows that the oscillator is doing well predicting event onsets, yet the correlation is not significant, indicating difficulty in point-to-point tracking. The figure shows how this can happen. The tempo curve for the second half of the performance, shown in panels (C) and (D), reveals timing deviations that strictly alternate: slower, faster, slower, faster. Panel (C) shows that the period tracking delta-rule is getting conflicting information, and the period curve remains basically flat. Panel (D) shows the effect of this pattern on observed cycle times. Cycle times are always one step behind the performed durations because changes to the oscillator's phase in the current cycle affect cycle time in the following cycle. In this case, timing deviations were "jagged," so when performed duration increased, cycle time decreased and vice-versa. Seven of the eight difficult cases match the profile of this case. Correlations between cycle times and target event IOI's were low, however average phase values were good showing that, as in this case, the oscillator successfully tracked the target event trains, doing a good job of predicting target event onsets.

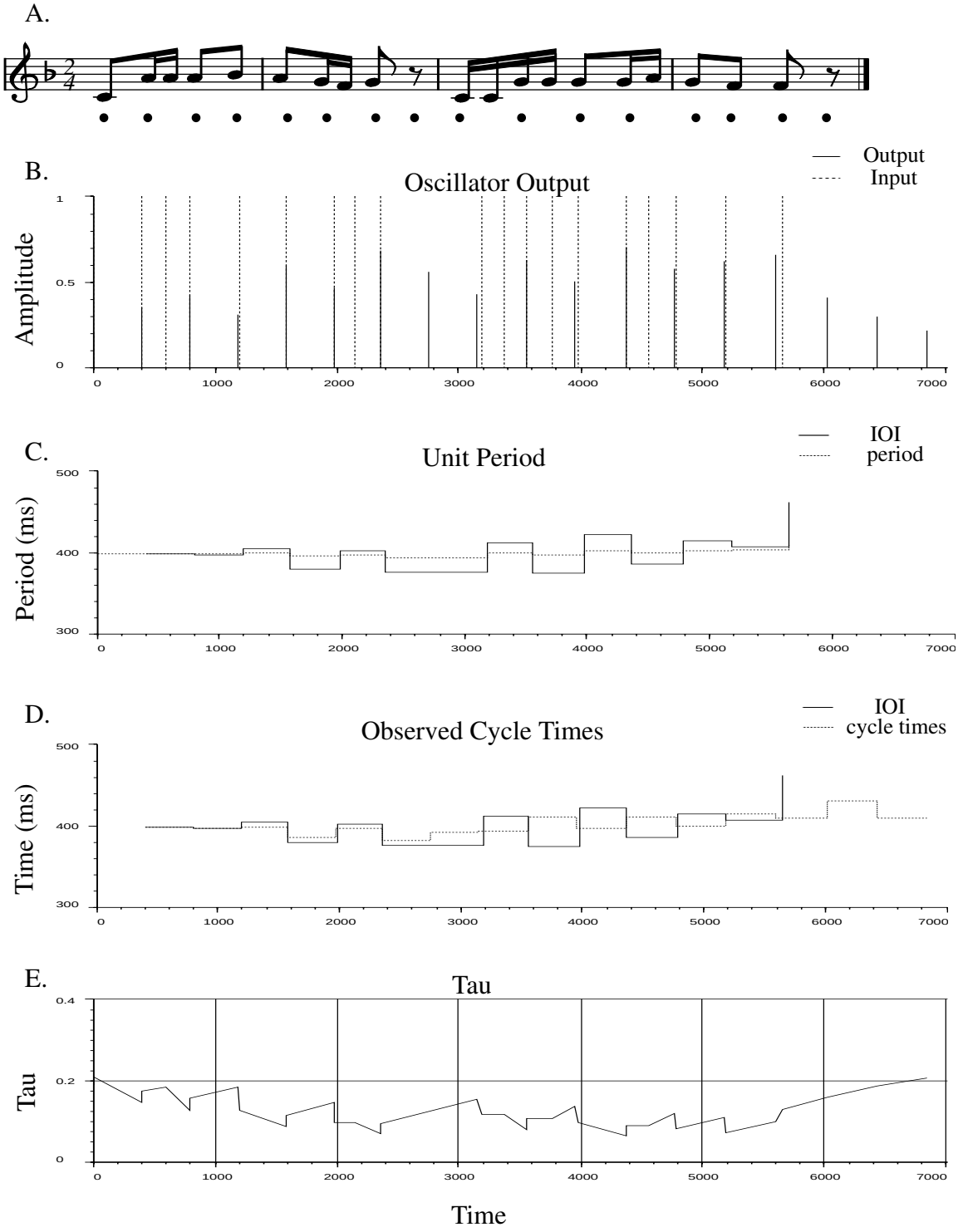


Figure 41: An oscillator tracking the rhythm of *Hush little baby* (rubato = 0.07, $\overline{|\phi'|} = 0.04$; $R^2 = 0.32$, $p = 0.29$).

8.1.1.2 Case 2

In 1 case out of 30 performances average phase was high and correlation was low, (rubato = 0.07, $\overline{|\phi'|} = 0.20$; $R^2 = 0.39$; $p = 0.18$), indicating true difficulty in tracking the target events. The largest value that $|\phi'|$ can assume is 0.5, when a target event onset occurs at anti-phase with the oscillator's pulses. Thus, this high value of average phase indicates that the unit had true difficulty tracking the target event train. Figure 42 shows what the difficulty was. At the end of the first half of the melody, the oscillator has done a good job of estimating period (panel (C)), but the first note of the second half comes in at nearly anti-phase from the oscillator pulses. For the next three cycles, the oscillator preserves this anti-phase relationship with the target event train; this is the source of the high value for average phase. On the third cycle, the oscillator performs a perceptual shift (a sudden 180 degree change in phase, as described in Chapter VI). The perceptual shift can be seen as a dramatic drop in observed cycle time in panel (D). After the shift, the oscillator successfully tracks the remainder of the performance.

Timing deviations in which the performer enters out of phase from the output pulses present a difficulty for the oscillator. The difficulty arises because target events occur at times when the oscillator has decided to ignore event onsets. However, the unit is able to recover because it can perform a perceptual shift. The value of τ rises (due to Equation 8, also due to decay of Ω), accompanied by a drop in confidence, allowing the shadow unit to take over (see Section 6.3.1). This presents a difficult situation for the oscillator, but one from which it can quickly and gracefully recover.

8.1.2 Performance of Improvised Variations

Next, the oscillator's performance in tracking the improvised variations was examined. The improvisations provided a more difficult situation than the performances of notated melodies for two reasons. The rhythms of the improvisations were more complex than the rhythms of the melodies, and the improvisations showed significantly greater timing deviations than did the performed melodies. The oscillator was exposed to thirty improvisations collected in the initial study – five improvised variations on three melodies by two pianists. An analysis of variance (ANOVA) was conducted with factors melody, pianist, and analysis type (rubato vs. phase). The ANOVA showed no main effect of analysis type ($F(1, 4) = 0.005$, $p = 0.947$), with mean rubato = 0.10, and average phase = 0.10. This result shows that for these performances, the oscillator does as well as the baseline strategy. Thus, on average the oscillator is tracking the target event trains well. The ANOVA also indicated a significant interaction of melody and subject ($F(2, 8) = 4.0$, $p < 0.05$).

To assess the performance of the oscillator more carefully, observed oscillator cycle times were compared with performed IOI's. Correlations between oscillator cycle times and performed IOI's were significant ($p < 0.05$) for 13 out of the 30 melodies. Out of the 17 nonsignificant results, seven melodies resembled Case 1 from the previous study, in which there was some lag between observed cycle times and performed IOI's, yet the oscillator tracked its target well ($\overline{|\phi'|} < 0.10$). This left 10 cases in which the unit had true difficulty in tracking its targets. These cases fell into three groups, Pianist 1's improvisations on *Mary had a little lamb* (Case 3), Pianist 2's improvisations on *Hush little baby* (Case 4), and Pianist 2's improvisations on *Baa baa black sheep* (Case 5).

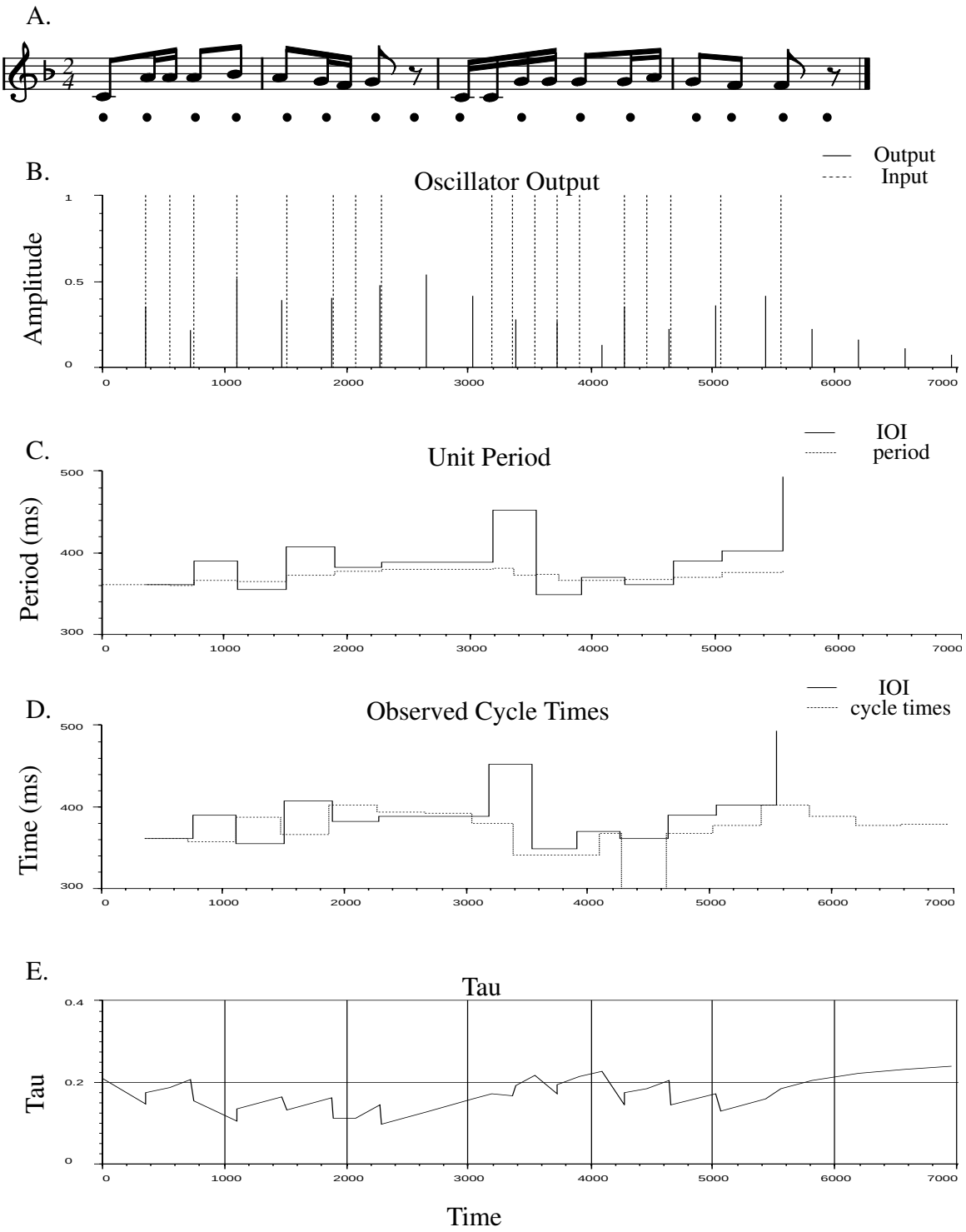


Figure 42: An oscillator tracking the rhythm of *Hush little baby* (rubato = 0.07, $\overline{|\phi'|}$ = 0.20; R^2 = 0.39; p = 0.18).

8.1.2.1 Case 3

Pianist 1's improvisations on *Mary had a little lamb* were performed in a freely timed blues style. The first improvisation had the highest rubato score (rubato = 0.25), highest average phase ($\overline{|\phi'|} = 0.17$), and the second worst correlation (R^2=0.04, p = 0.87). The oscillator's behavior in this case was representative of its performance on this group of melodies, so it was chosen for further study. The time series corresponding to the performance of the oscillator are shown in Figure 43. The tempo curve indicates the presence of large timing deviations at several points in the melody. Points of particular difficulty are around t = 3000*ms*, t= 8000*ms*, t= 16000*ms*, t= 18000*ms*, and t = 23000*ms*. At these points, τ rises (correspondingly, confidence drops) allowing the oscillator to continue to track the target in spite of the large deviations. In three of these cases the compound unit responds to difficult timings with perceptual shifts.

In spite of these difficulties, however, the figure shows that the oscillator did a good job of tracking its target event train in this rhythm. Beats are output at approximately the correct times throughout the piece – the oscillator is not lured away by the many distractor events in this rhythmically complex performance. Another way to see this is to note that the value of average phase ($\overline{|\phi'|} = 0.17$) is lower than the rubato measure (0.25). Additionally, oscillator confidence is high for large sections of the piece; by the oscillator's internal measure its performance is good.

8.1.2.2 Case 4

The next challenge was posed by Pianist 2' improvisations on *Hush little baby.* These improvisations were the most varied of all the improvisations studied here and made heavy use of rubato. The improvisation that proved the most difficult for the model to handle, the third variation, was chosen for further study. This improvisation made the greatest use of rubato and had the highest average phase, (rubato = 0.16), $\overline{|\phi'|} = 0.30$) and the second worst correlation (R^2=0.04, p = 0.22). These numbers suggest extreme difficulty in tracking. Figure 44 shows the actual time series corresponding to the performance of the oscillator. Throughout this improvisation the performer makes use of the sort of "jagged" rubato seen in Case 1, above. In the current case, however, the amount of rubato is so large as to pose a serious difficulty for the model. The event at time t = 1800*ms* is very early (almost anti-phase) the confidence of the unit decreases. By the downbeat of the second full measure, a perceptual shift takes place. This situation repeats itself until nearly the end of the performance. Toward the end the performer regulates the timing, and the unit finally begins to pick up confidence.

In spite of these difficulties, the oscillator did track this performance, but in an odd way. The units responds to the correct events, and outputs beats at approximately the correct locations throughout the piece. However, the oscillator's confidence is consistently low, because there is so much temporal deviation. This represents the limiting case for the model: beats are output in the more-or-less correct locations, but the oscillator's internal measure of performance is low. Thus, the performance is limited for large, "jagged" timing deviations.

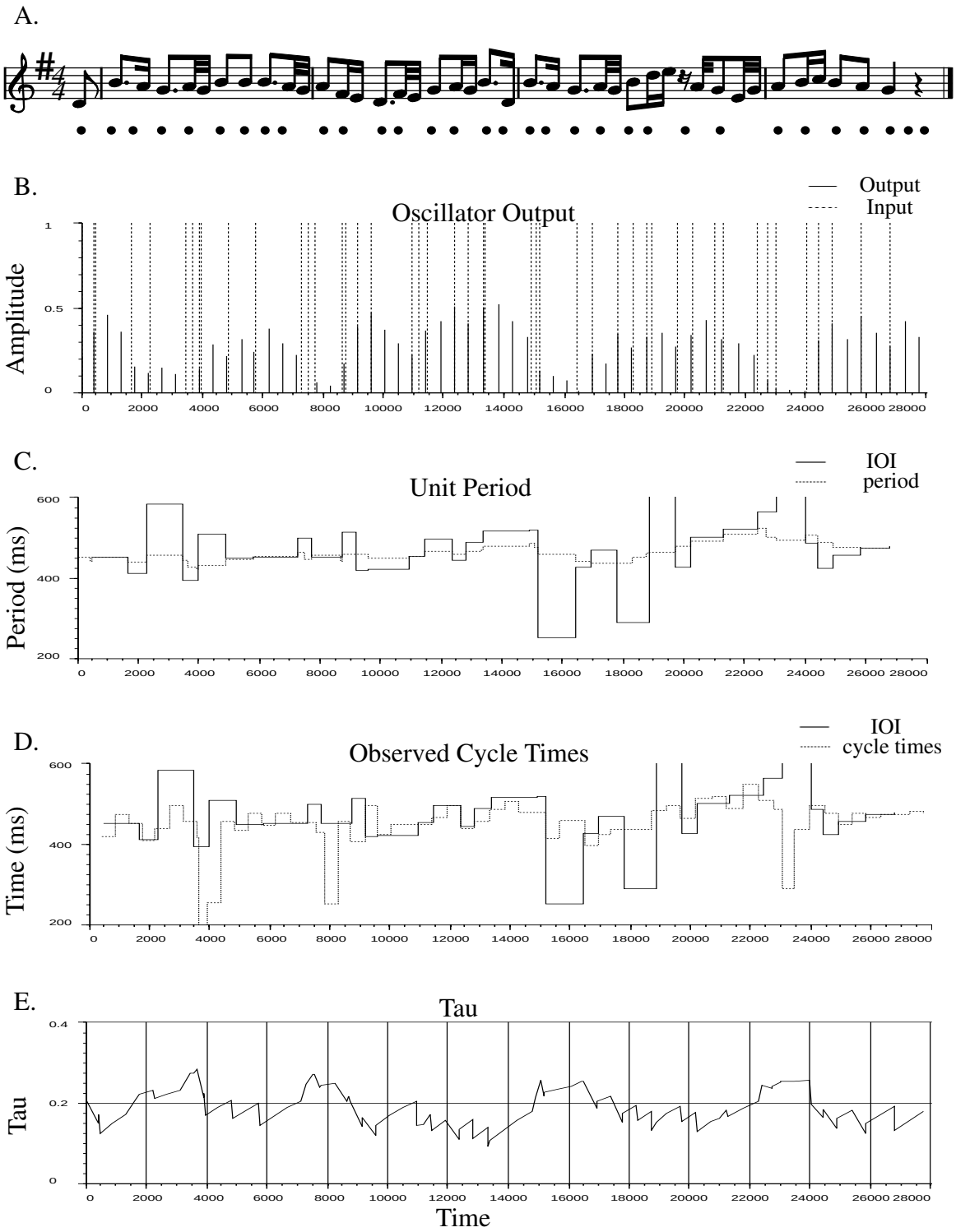


Figure 43: Oscillator tracking an improvisation on *Mary had a little lamb* (grace notes are not transcribed); rubato = 0.25, $\overline{|\phi'|} = 0.17$, R^2=0.04, p = 0.87).

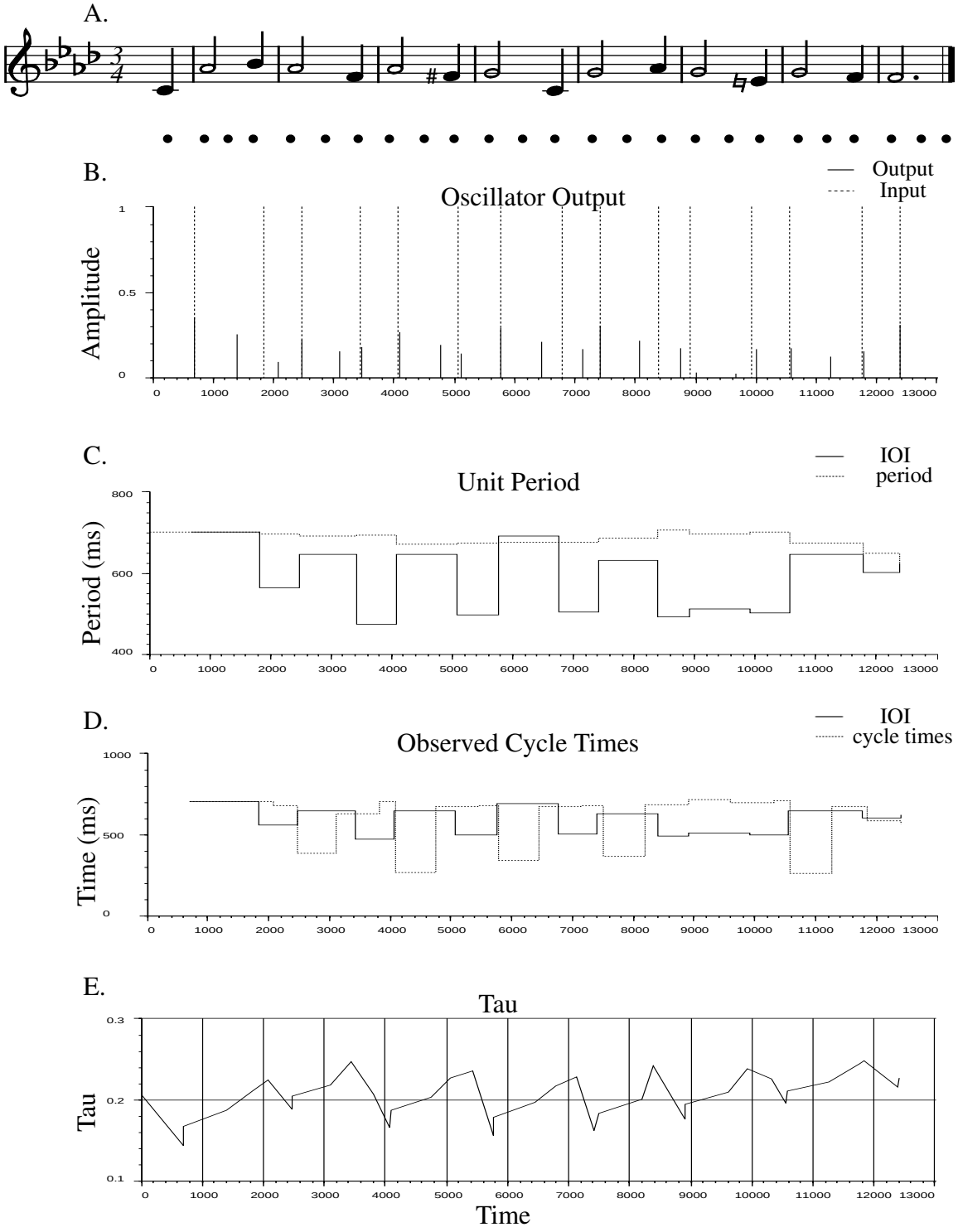


Figure 44: Oscillator tracking an improvisation on *Hush little baby* (rubato = 0.25, $\overline{|\phi'|} = 0.17$; R^2=0.04, p = 0.87).

8.1.2.3 Case 5

The final challenge to the model was posed by Pianist 2's improvisations on *Baa baa black sheep*. Three of these five improvisations have moderately low correlation scores. The most difficult case was given by the final improvisation. The final improvisation had largest timing deviations, the highest average phase and a moderately low correlation score (rubato = 0.39, $\overline{|\phi'|}$ = 0.20; R^2=0.20, p = 0.09). This improvisation is different from the others, however, because of the source of the timing deviations: timing errors. Timing errors were defined as situations in which the trained analysts were not confident in the transcriptions they prepared; they thought they were forced to make guesses about what the performers intended. According to this definition, performance errors occurred in three of the improvised variations. As discussed in Chapter III, these data were retained because they provide a valuable source of real-world noise in the test data.

The behavior of the model in response to this improvisation is shown in Figure 45. Because this performance is so long and complex, only the final two-thirds of the performance are shown, and no transcription was prepared for the figure. The large timing deviations between t=18000 and t=20000*ms*, between t=23000 and t=24000*ms*, and again between t=28000 and t=30000*ms* are timing errors. In the first and third errors, the performer appeared to stumble over the complex ornaments he was improvising; the second error was a pause.

In spite of these large deviations, the oscillator tracked this performance well. Average phase was high due to temporal deviations, but correlation approaches significance, indicating some success in point-to-point tracking. τ increases and confidence drops when deviations occur, but when the timing recovers the oscillator correctly picks up the target event train, and confidence grows again.

8.2 Discussion

These analyses suggest that the compound oscillator, coupled with complex, nonstationary rhythms that arise from musical performance, may adequately model the perception of musical beat. Melodies may be the most difficult case for an entrainment model, because they tend to provide fewer reliable cues to entrainment than accompanied melodies. In 49 out of 60 cases, the oscillator tracked performed rhythms well. In 11 cases, difficulties were encountered. These difficulties were caused by large temporal deviations, stemming from three sources: heavy use of rubato including 'phase-shifts' (Cases 2 and 3), jagged rubato curves from alternating shortened and lengthened durations (Cases 1 and 4), and actual timing errors (Case 5). In every case but one (Case 4) the oscillator still tracked adequately, outputting beats at appropriate times, although sometimes with temporarily lowered confidence. In Case 4, the oscillator output beats at approximately correct times, but with consistently lowered confidence. Jagged rubato forced the unit to maintain a high value of τ to track the performance, and because confidence was defined to have an inverse relationship with τ, the unit was not able to recognize that it was outputting beats at the correct times. Overall, the oscillator performed very well in tracking complex melodies with no information other than event onset times.

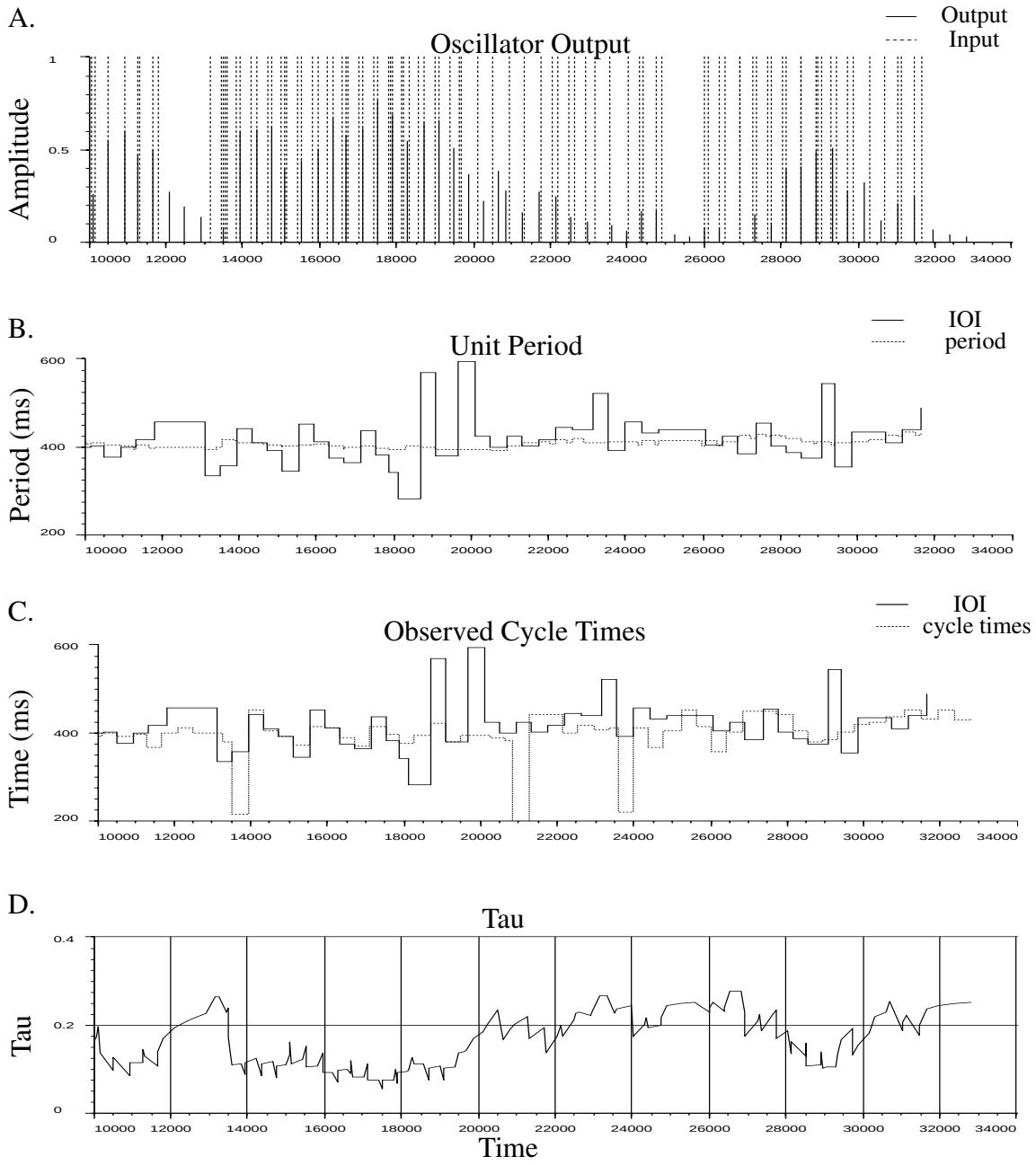


Figure 45: Oscillator tracking an improvisation on *Baa baa black sheep* (not transcribed; rubato = 0.39, $\overline{|\phi'|} = 0.20$; R^2=0.20, p = 0.09).

8.3 The Perception of Metric Relationships

The next experiment investigated the usefulness of the oscillator in modeling the perception of metrical relationships. The response of a system of oscillators to stationary rhythms of varying levels of structural complexity (hierarchical and polyrhythmic metrical structures) was examined using four input signals. Each signal was composed of two isochronous event trains. Input amplitude corresponding to each event was arbitrarily set to 1, and signals corresponding to the two events trains were summed. Thus, when input events co-occurred amplitude was equal to 2. Figure 46 shows the four test signals in schematic.

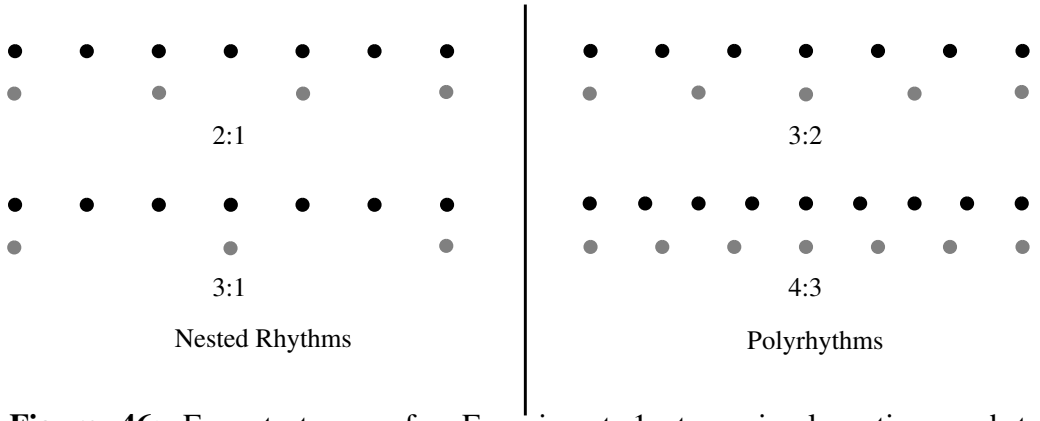


Figure 46: Four test case for Experiment 1: two simple ratios, and two polyrhythmic ratios.

A system of oscillators with different resting periods was used to identify metrical relationships. Such systems are useful for self-organizing metrical responses to rhythmic stimuli (Large & Kolen, in press). The resting periods of the oscillators in the system were defined such that the system spanned two octaves with two oscillators per octave according to the relationship:

$$p_{i+1} = 2^{\sqrt[3]{p_i}}.$$

Next, each oscillator's period range was limited by allowing the period of each oscillator to decay back toward its resting period. The resting periods of the four oscillators were 240*ms*, 339*ms*, 480*ms*, 720*ms*, respectively and the following parameters were used for each oscillator $\eta_1 = 1.0$, $\eta_2 = 0.2$, $\eta_3 = 0.3$. Each oscillator responded independently to the signal. At time $t = 0$ each oscillator had $\phi = 0$.

The first rhythm was constructed using an inter-onset duration of 240*ms* for the first event train and 480*ms* for the second event train, giving a simple 2:1 rhythm. Figure 47 shows the response of Units 1 and 3 (the two oscillators with the appropriate period ranges) for this rhythm. As shown in the figure, the units' response pattern mirrors the 2:1 pattern.

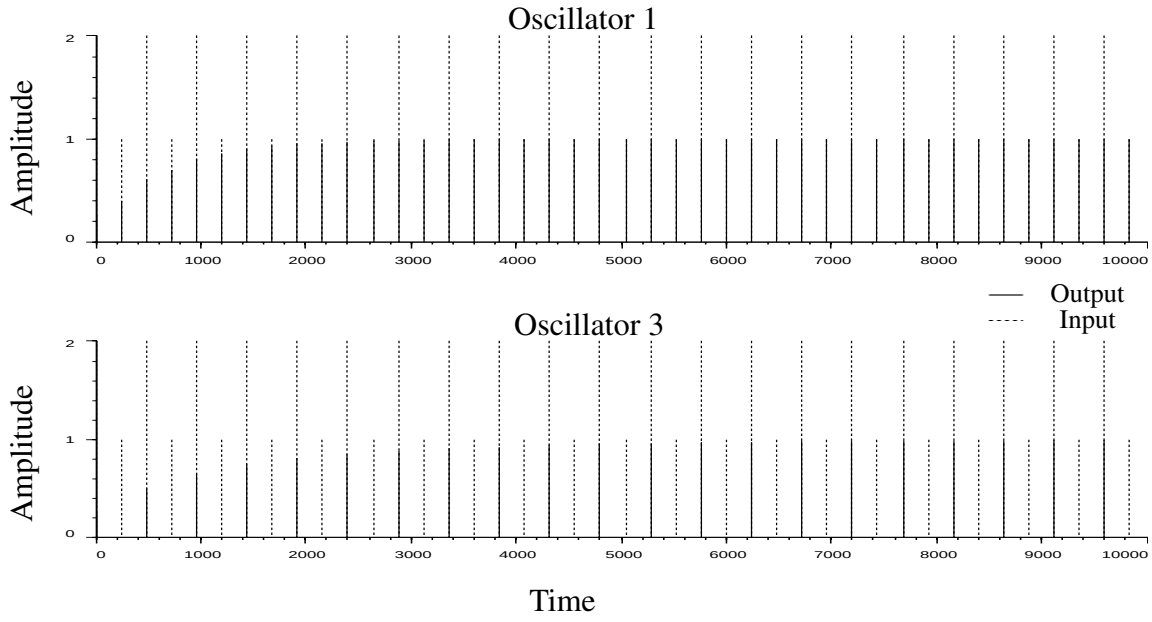


Figure 47: Response of Oscillators 1 and 3 to a 2:1 rhythm.

Next, a 3:1 pattern was constructed using inter-onset durations of 240*ms* and 720*ms*. In this case Units 1 and 4 have the appropriate period ranges, and their response is shown in Figure 48.

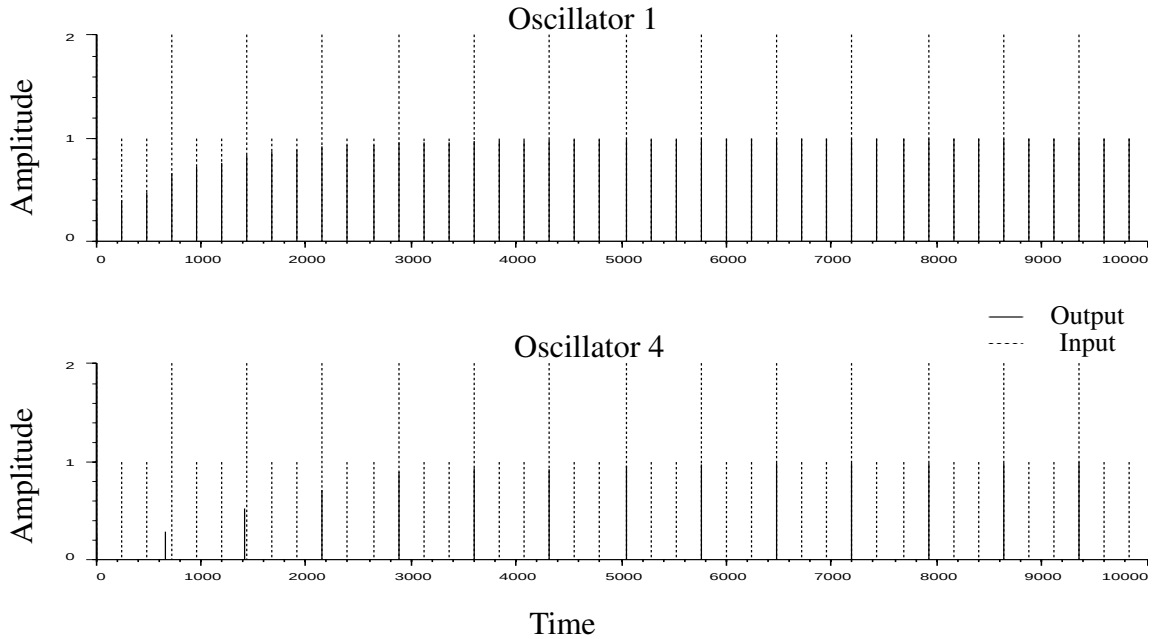


Figure 48: Response of Oscillators 1 and 4 to a 3:1 rhythm.

Next polyrhythms were tested. The first polyrhythm was a 3:2 pattern constructed using inter-onset durations of 240*ms* and 360*ms*. Figure 49 shows the response of Units 1 and 2.

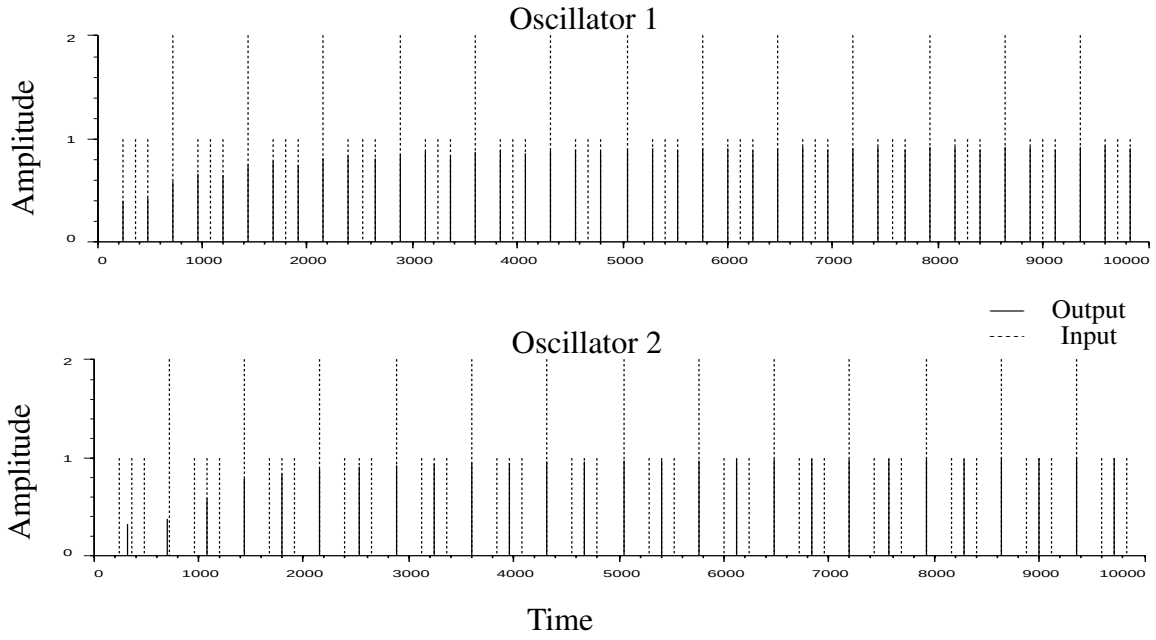


Figure 49: Response of Oscillators 1 and 2 to a 3:2 polyrhythm.

Finally, a 4:3 polyrhythm was constructed using inter-onset durations of 240ms and 300ms. Figure 50 shows the response of Units 1 and 2, again displaying the appropriate phase and period relationships.

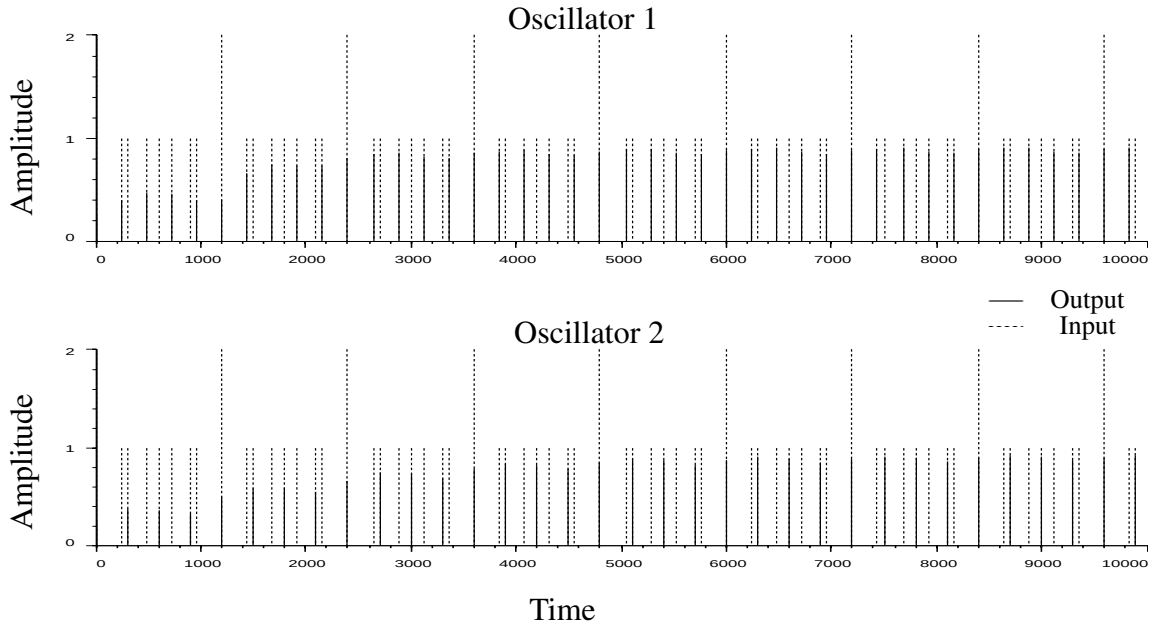


Figure 50: Response of Oscillators 1 and 2 to a 4:3 polyrhythm.

Figure 50 shows output strength for each oscillator, for each rhythm. This figure shows that oscillators with inappropriate period ranges have lowered output confidence. This simple system of units has correctly parsed these four structural relationships. The figure also shows that the more complex the rhythm, the more difficult the lock is to acquire between the rhythm and the responding oscillators.

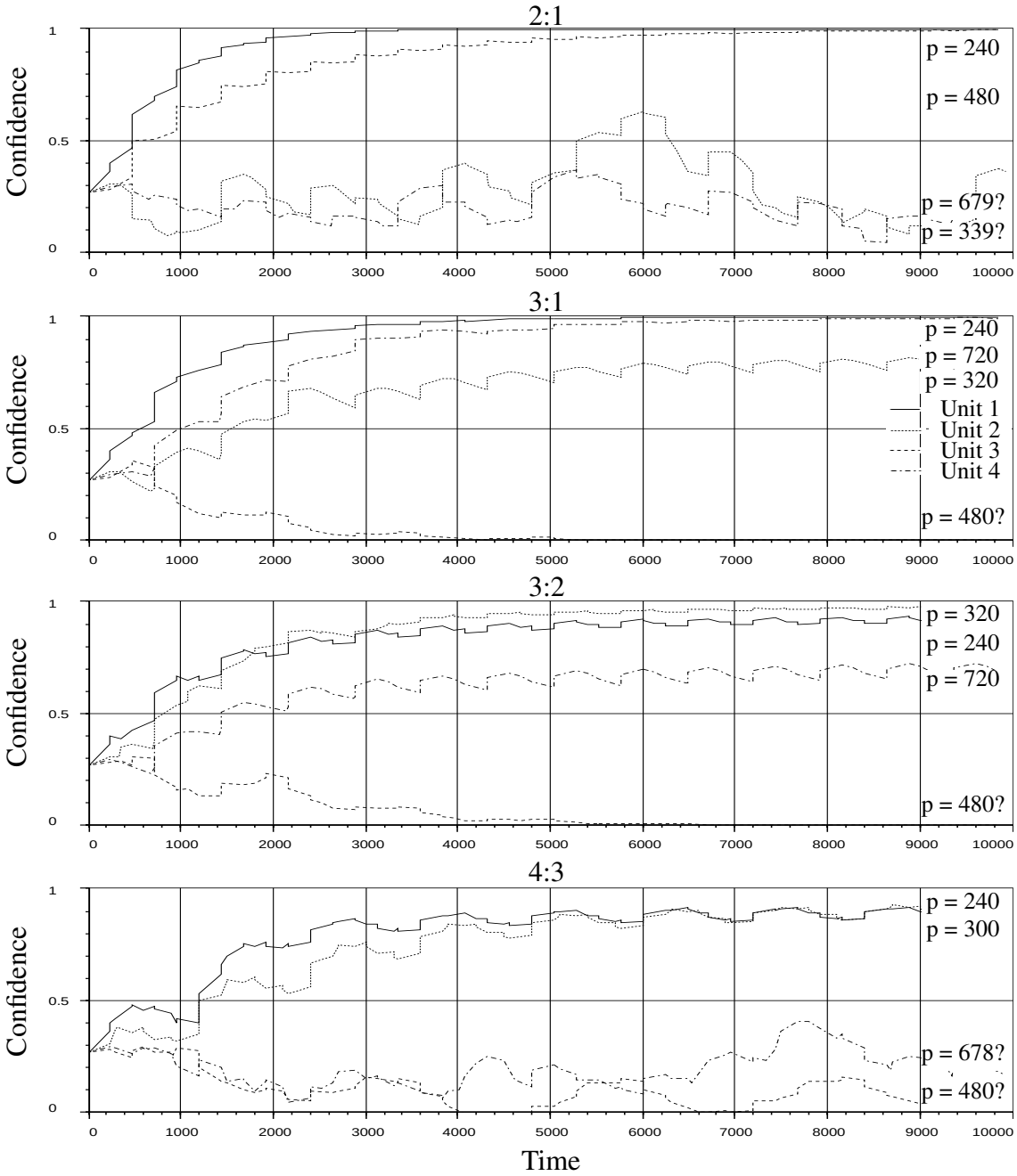


Figure 51: Confidence of each oscillator and rhythm. Each curve is marked with the period of the corresponding oscillator. Those that do not acquire stable mode locks are indicated with '?'.

The results of this study confirm the implications of the regime diagram analysis of Chapter VII. Appropriately constructed networks of oscillators can parse metrical structures, not only for simple hierarchically nested structures, but also for more complex polyrhythmic structures. The response of the system also suggests a type of perceptual prediction made by the model: that more complex polyrhythms are less perceptually stable. For example, this model predicts that time discrimination for polyrhythms may be more difficult than for simple hierarchically structured rhythms (cf. Yee, Holleran, & Jones, in press).

8.4 Discussion

Previous chapters introduced and developed a dynamical system model of beat perception. The goal of this chapter has been to determine the suitability of the model for explaining beat perception and meter perception in complex musical rhythms. Three types of complexity found in music performances provided challenges to oscillator performance: rhythmic complexity, timing deviation, and structural complexity. Two experiments assessed the ability of the model to cope with these difficulties.

The first study addressed rhythmic complexity and timing deviation in music performance at two levels of difficulty: performances of simple melodies and more temporally complex melodic improvisations. Melodies (single musical voices) are a difficult case for the entrainment model, because they usually provide fewer cues for entrainment than accompanied melodies. At both levels of difficulty model performed well, tracking beats with little difficulty. These results are interesting because signal impulses did not carry amplitude (accent) information; the markers of the oscillator's target event train were not distinguished from other impulses. Thus, entrainment can occur even when phenomenal accent information is missing, ambiguous, or misleading.

The second study dealt with the issue of structural complexity. A network of four units was exposed to stationary rhythms of varying structural complexity, 2:1, 3:1, 3:2, and 4:3. The network demonstrated appropriate behavior for each rhythm. Co-occurrence of events was represented in the input using amplitude information. Oscillators whose range contained a periodicity of the input rhythm tracked the corresponding event train, and confidence increased. Oscillators whose range did not include a periodicity of the input rhythm decreased in confidence, turning themselves down or off. The system of oscillators demonstrated the ability to entrain both to simple harmonic rhythms and to dissonant polyrhythms. Entrainment modes corresponding polyrhythms were marked by longer-lasting transients and lower oscillator confidence once entrainment was achieved. When amplitude information is available, an oscillator may use it to boost its confidence. Such behavior can be exploited to allow a system of units to parse the meter of complex signal. Thus the oscillator can use amplitude information when it is available, although it does not depend upon it for entrainment.

One difficulty arises in cases in which very large timing deviations require high values of τ so that the oscillator may track the target events adequately. The difficulty arises because confidence, c, is inversely related to τ, so the oscillator's internal measure of performance is low. In certain musical situations the resulting behavior may be inappropriate. However, the

relationship between τ and c seems appropriate given the restricted context of the model (temporal relationships). Perhaps an oscillator such as this, embedded in a larger temporal sequence processing context could contribute to appropriate behavior. If the oscillator were coupled with a network making predictions about what events were to occur, for example, performance at the sequence prediction task could be used to boost the system's internal measure of performance. This could be achieved by coupling this oscillator with a strategy for on-line adjustment of processing parameters for learned sequences (e.g. Cottrell, Nguyen, & Tsung, 1993). The combined system would show greater tolerance for timing deviations in learned compared to unlearned sequences. This represents a more complex strategy than simply using amplitude information, but may be useful in certain contexts.

In summary, these results strongly support this dynamical system model as one that may have sufficient power to explain the perception of beat and meter in music performance. The model can parse simple metrical structures, as well as polyrhythmic structures. In addition, it can handle rhythmic complexities and temporal deviations associated with musical performance and improvisation. The model does not fall into the trap of phase-locked loop models, in which a subset of events must first be marked as accented or stressed before entrainment can take place. The oscillator can flexibly incorporate additional information (e.g. accent) to parse metrical structures, but it does not depend on such information for its basic operation. Thus, this model may be suitable for use in combination with other temporal sequence processing strategies to provide temporal structure information that can be used to bootstrap learning of complex, temporally structured sequences.

CHAPTER IX

IMPLICATIONS: MUSIC COGNITION AND BEYOND

The goal of this dissertation has been to understand how complex, temporally structured sequences may be coded as patterns of activation in artificial neural networks. The domain of the studies reported here was music, and simulation results were evaluated with data from skilled music performance. Two issues were addressed. The first was the acquisition and representation of structural relationships among sequence events. A model of sequence coding was proposed that captured relative importance among sequence elements. The second was the representation of temporal relationships among events. A model of entrainment was proposed, from which a dynamical system model of beat perception and a preliminary model of meter perception were developed.

9.1 The Basic Findings

9.1.1 Computing Structural Descriptions for Musical Sequences

The first model computed reduced distributed representations that captured a specific type of relationship among sequence elements, the relative importance of events. Phenomena such as style acquisition and the recognition of musical variation can be explained by positing such mechanisms. The reduced memory descriptions computed for melodies resulted from encoding and decoding mechanisms that compressed and reconstructed the musical sequences. These mechanisms led to reduced descriptions similar to those predicted by reductionist music theories. This type of memory representation abstracts and summarizes sections of musical material, extracting what Dowling and Harwood (1986) call the "gist" of a musical sequence. The reduced representations are suitable for manipulation by other neural-style processing mechanisms, and therefore may be useful for modeling musical activities such as expectation, improvisation, and sequence recognition. A general learning algorithm (backpropagation) provided an example of how the knowledge for computing reduced memory descriptions may be extracted from a learning environment, addressing an important challenge to reductionist theories. These findings support reductionist theories of music comprehension, suggesting that the computation of musical reduction may not be an end in itself; rather, it is a natural result of the construction of memory representations for musical sequences.

Agreement was found among evidence from improvisational music performance, the neural network model, and theoretical predictions regarding the relative importance of musical

events. These findings support the psychological plausibility of reductionist theories of music comprehension. The fact that musical events were weighted similarly in musicians' choices of events to retain in improvisations, network encodings of the same melodies, and theoretical predictions of relative importance suggests that recursive distributed representations capture relevant properties of humans' mental representations for musical melodies.

9.1.2 Dynamic Representation of Temporal Structure

The model of the dynamic representation of temporal structure was based on a simplification of the dynamic attending hypothesis (Jones, 1976), called the *entrainment hypothesis*. The entrainment hypothesis proposes that a basic mechanism of time perception is entrainment of perceptual processes to pseudo-periodic components of rhythmic patterns. An oscillator was proposed that synchronizes its behavior to rhythmic patterns. The oscillator responds to event onsets that occur within a temporal receptive field, and ignores stimulus pulses that occur outside this field, enabling it to isolate pseudo-periodic components of complex rhythms. A number of such processes may be composed to reverse-engineer the structure of motor programs to reconstruct perceived rhythms.

Based upon the oscillator model, a dynamical system was constructed to model beat perception, revealing complex dynamics. The oscillator can mode-lock to a periodic stimulus in any one of an infinite number of rational ratios. Tuning of the oscillator's temporal receptive field has the effect of adjusting the relative stability of various mode-locking regions. Large temporal receptive fields result in a preference for simple ratios. Finely tuned regions allow more complex ratios. These properties have important implications for any theory of temporal structure that includes entrainment as a primary component. Regime diagrams summarize the content of an entrainment theory regarding the well-formedness of temporal structures. Simulation of oscillator responses to complex polyrhythmic ratios, and to rhythmically complex, nonstationary rhythms derived from musical performances showed the robustness of the entrainment approach to meter perception. Unlike previous theoretical and computational approaches, this model is well-suited to handling rhythmic complexity, timing deviation, and structural complexity in real-time. The success of the model, as tested so far, provides strong support for models of meter perception and temporal expectancy that implicate entrainment as the basic mechanism for the perception of temporal structure.

9.1.3 Sequence Structure and Temporal Structure

The RAAM network model of sequence representation used knowledge of temporal structure to efficiently encode of sequence structure. The network's chunk-and-recode strategy allowed it to successfully represent the long, complex sequences found in melodies. In addition, reduced descriptions produced in this way captured an important form of structural relationship among sequence elements, the relative importance of events. Knowledge of temporal structure was exploited by the network to extract stylistic regularities that are systematically related to relative timing relationships. The network learned relative importance because each position in its input buffer corresponded to a metrical grid location, and the network used a dedicated set of weights for

each position. This strategy made the network sensitive to relative timing relationships in a unique way. The entrainment model provided a way to identify such structure in complex temporal sequences.

9.2 General Significance

9.2.1 Connectionist Temporal Sequence Processing

The entrainment model of beat perception focused exclusively on temporal processing issues, yet it was motivated from the point of view of temporal sequence processing. This model, interpreted in light of the results of the neural network simulation, suggests a way to build rate invariant temporal sequence processing networks. Current connectionist approaches address this issue by using a system of short term memory delays to explicitly capture temporal context (Lang, Waibel, & Hinton, 1990; Unnikrishnan, Hopfield & Tank, 1991; Bodenhausen & Waibel, 1991; de Vries & Principe, 1992). Delays may be hardwired or learned during batch training, but during processing they remain fixed. The problem with a fixed memory delay solution for music processing, however, is that timing in music is too flexible. Musical signals display large, systematic deviation from timing regularity. Cottrell, Nguyen, and Tsung (1993) have addressed this problem using a recurrent network that controls its own processing rate by adapting time constants and processing delays. The drawback of this approach is that it applies only to learned sequences.

The oscillator responded flexibly and on-line to changes in presentation rate without needing to memorize sequences in advance. This behavior was enabled by the fact that the oscillator ignored the sequence content (events), dealing only with rhythm. Therefore, this strategy is applicable both to learned and unlearned sequences. Such behavior might be applied to adjust of neural network short term memory parameters on-line. Memory delays implemented as resonance-based components, for example, could adapt to the rate of rhythmic signals by tracking pseudo-periodic components of rhythmic sequences. Using several oscillators, the actual structure of short term memory could adapt to reflect the temporal organization (e.g. the metrical structure) of an incoming signal. Thus, entrainment-based memories provide a novel approach to the problem of rate-invariance in temporal sequence processing.

9.2.1.1 Oscillation and Synchronization in Dynamic Feature Binding Networks

Phase-locking phenomena have been of interest in the connectionist community for some time, especially since the discovery of oscillations and synchronization behavior in the cat visual cortex (Eckhorn, et. al., 1989; Gray, et. al., 1989). It has been proposed that the oscillations of neurons in the cat visual cortex phase-lock to establish relations between features in different parts of the visual field (Gray, et. al., 1989). It has further been suggested that the brain could be using synchronized oscillations as a general method of solving the binding problem (von der Malsberg, & Schneider, 1986). Phase-locking may add an extra degree of freedom to neural network models, so that several different entities may be represented simultaneously using the same set of units, each by a different phase in an oscillatory cycle.

The use of oscillation reported here differs from that proposed in the literature on neural feature binding. First, work on neural feature binding focuses mainly on synchronization among a population of oscillators (Wang, in press b). The current work focuses upon the synchronization of internal perceptual processes with complex external rhythms. Second, rather than using coupled oscillations to describe a neural strategy for performing an implementation-level operation such as feature binding, the entrainment model used synchronization to describe how the brain may execute the relatively high-level cognitive functions of beat and meter perception.

The neural feature binding literature may offer insight into neural implementation of the proposed mechanism, however. For example, a pair of units could produce the oscillatory behavior of interest (e.g. Wang, 1993). McAuley (1993; 1994), who has also proposed entrainment models for the perception of rhythm, has suggested that behavior relevant to this task, including period-tracking, may be found at the single neuron level. These possibilities are intriguing; however, I have proposed a functional approach, not an implementation-level strategy. Therefore I assume that the abstract, functional oscillators represent higher levels of abstraction than individual neurons. The behavior of the abstract oscillators may be plausibly regarded as the emergent behavior of a wide range of possible brain structures from simple neuronal substructures to large networks of oscillatory neurons.

9.2.2 Speech

The above discussion raises the possibility that the mechanisms proposed for music processing may be applied to the processing of other complex signals such as speech. Rhythmicity has been difficult to identify in natural language, but appears to exist perceptually (Lehiste, 1977). There are two complexities in musical rhythm that bear upon this problem. The first is the temporal deviation in music performance; the components of musical rhythms change frequency. Thus strict timing does not usually exist in music either. The second is rhythmic complexity, including syncopation. Ideally phenomenal accents in music correspond to strong metrical locations, but interesting music is not usually composed in this way. Stressed events are placed in weak metrical locations to create interest.

Perceptual isochrony in language may be modeled much as beat perception in music. Perhaps it could be modeled by the same dynamical system proposed here for beat perception in music. As a preliminary test of this hypothesis two oscillators were exposed to a time-series of impulses from a digitally sampled acoustic signal of read speech. The signal was reduced to a series of impulses using a one-dimensional version of an edge-detection algorithm that is common in computer vision applications (Marr & Hildreth, 1980, N. Todd, 1994). The edge detection algorithm found the onset of events that corresponded to syllables, marking some, but not all of the syllables in the spoken excerpt. The algorithm identified mainly stressed syllables, but there were exceptions.

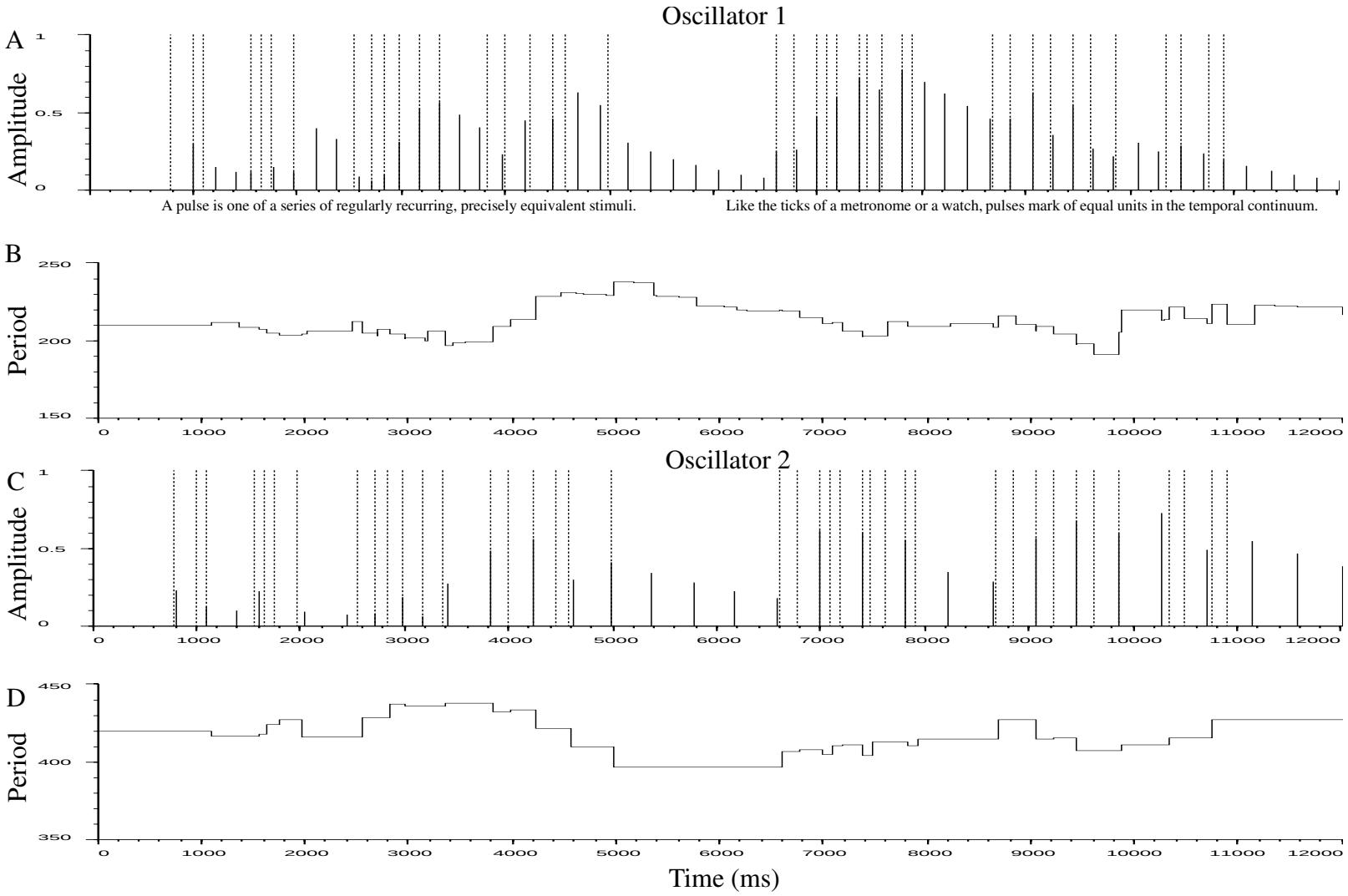


Figure 52: The response of two oscillators to rhythmic input from a digital sample of read speech (from Cooper & Meyer, 1960). The digital sample was filtered using an edge-detection algorithm, and the results of the edge detection algorithm were used as input. The input impulses (dotted lines) and output beats (solid lines) are shown for Oscillators 1 and 2 in panels A and C, respectively. The periods of Oscillators 1 and 2 are shown in panels B and D, respectively.

Figure 52 shows the response of the oscillators to this input. Output amplitudes show that the oscillators tracked two pseudo-periodic components confidently. The output pulses provide the type of information necessary for time-warping in speech recognition. They also provide information about systematic timing deviations, useful for syntactic and semantic processing of the speech signal (Lehiste, 1977; Cutler & Mehler, 1992). When one listens to this example, one is struck by the complex polyrhythms that can be heard against the taps provided as the outputs of the oscillators. Although this little more than anecdotal evidence, it presents intriguing possibilities for future work. This dynamical system model may provide a novel method of measuring the periodic structure present in the rhythm of speech.

9.2.3 Motor Coordination

Many activities, including rhythmic hand movements, cascade juggling, and piano performance are consistent with mathematical laws governing coupled oscillations (e.g. Kelso & deGuzman, 1988; Schmidt et. al., 1991; Treffner, & Turvey, 1993; Shaffer, 1981; for a review of models see Beek, Peper, & van Wieringen, 1992). Studies of motor control often assume (even rely on the fact) that subject are able to synchronize with external signals (e.g. Kelso & deGuzman, 1988), but mechanisms for learning and adaptation have been less widely studied (but see Zanzone & Kelso, 1992). Thus, one difficulty with this approach to motor control lies in explaining how perceptual systems cope with rhythmic complexities in entraining to rhythmic signals.

The entrainment model addresses this problem. This approach treats the object of meter perception as a motor program. An analysis of coupled oscillation uncovered mathematical constraints on the perceptual model that are consistent with the motor constraints posited in the coupled oscillator approach to motor control (Kelso & deGuzman, 1988; Treffner, & Turvey, 1993; Beek, Peper, & van Wieringen, 1992). The perceptual system may cope with rhythmic complexities by adjusting entrainment parameters that, in effect, adjust the size and stability of mode locking regions seen in regime diagrams (Chapter VII). These perceptual constraints provide an efficient way to perform the task of reverse engineering motor programs through self-organization. Thus this perceptual work supports coupled oscillator theories of motor coordination.

9.3 Future Work

The approach to understanding structural descriptions for musical sequences presented here has some limitations that highlight the need for further work. One regards the choice of musical materials in the empirical studies of Chapter III. For one thing, the use of musical materials as simple as these melodies leads to some difficulties in interpreting the network findings. It is not clear whether the network's representational capability at global structural levels was limited by the network architecture or by the choice of training materials. In addition the relationship between metrical accent and time-span reduction predictions of importance were not controlled; the restriction to a small set of musical materials makes it difficult to determine how the network learns relative importance independently of metrical accent or how one might model these structural relationships in more complex forms of music. Thus, it is difficult to say precisely what structural

relationships the RAAM model is capable of learning. Further study might use training and test melodies that control for interactions among structural relationships (cf. de Vries & Principe, 1992).

Another possibility for further work concerns the design of neural network architectures. One constraint of the RAAM architecture is the requirement of an external stack control mechanism for handling intermediate results during encoding and decoding (Pollack, 1988, 1990). In addition, the model requires a fixed-structure input buffer to exploit temporal information, such as metrical structure. The entrainment model presents the possibility that more flexible, self-organizing short term memory structures may be designed to make use of temporal structure in sequence processing. One possibility is that recurrent network architectures might be adapted to exploit temporal information in a more flexible way using entrainment based short term memory designs.

A primary goal of the oscillator model was to understand the implications of entrainment for the perception of metrical structure. The oscillator's ability to entrain to signals regardless of the nature of events was emphasized. However, accent in music is important for the perception of meter. The second experiment of Chapter VII suggested one way accent could be incorporated into the model: as the amplitude of input pulses. Amplitude may be used to increase oscillator confidence, allowing a group of units to parse metrical structures. This is an interesting possibility; however, use of such a strategy would require a theory of phenomenal accent. One possibility N. Todd's (1994) rhythmogram model. Using a direct temporal analog to the ∇^2 operator for spatial edge detection (Marr & Hildreth, 1980; Marr, 1982) Todd's system simultaneously carries out auditory edge detection at multiple time scales. The analysis yields onsets times and accent strength, precisely the type of information needed to drive the oscillator model.

The second issue for the perception of musical meter is that of network construction. I have concentrated on the behavior of individual units responding independently to rhythmic patterns. One would expect individual units within a network to interact, responding to the outputs of other units in the network. The important question is: Could a stable response emerge from such a network subjected to a musical event sequence? Analysis of the single oscillator case suggests that subsets of units in a loosely coupled network could self-organize a coherent response to a rhythmic input signal. The interaction would instantiate metrical well-formedness constraints. The challenge facing this approach is to determine the nature of the interactions. I leave the issue of network construction unresolved, and regard this as an important area for future exploration.

Finally, discussion of the entrainment model emphasized applications to rate-invariant temporal sequence processing. In a sequence processing context oscillator performance may improve. In some difficult situations, the oscillator was able to track the rhythm of the input signal, but variability was so high that oscillator confidence (the oscillator's internal measure of performance) was low. Within a sequence processing framework, predictions of *what* events are to occur might also be used to influence confidence. Successful prediction of what is to occur next could help make up for high variability in performance timing. A strategy such as Cottrell, Nguyen, and Tsung's (1993), of adapting processing rate according performance at sequence prediction, might be incorporated into delta rules to affect phase and period adjustments. In this situation,

learning would proceed by a kind of bootstrapping: in an unfamiliar domain the oscillator would control the sequence processing network, and sequences would be learned in conditions of normal timing deviations. As knowledge of sequence structure was acquired, networks would learn to tolerate even very large deviations for known sequences.

9.4 Closing Thoughts

Lashley (1951) identified the problem of serial order, "... the logical and orderly arrangement of thought and action," as a central problem for those who ultimately wish to describe the phenomena of mind in terms of the mathematical and physical sciences. Lashley realized that the problem was not merely one of sequence processing. The temporal structure of human perception and action implies that the temporal structure of neural computation is extraordinarily complex (Lashley, 1951). In this regard, the study of music is invaluable to the understanding of neural computation. Music forces us to deal with all aspects of time: time is so fundamental to music that it cannot be conveniently and convincingly abstracted away. It may even be that composers and performers shape the temporal structure of music to reflect and to explore natural modes of temporal organization in the human nervous system.

For inherently temporal tasks, such as perception and motor coordination, resonance may provide a more useful metaphor than general-purpose computation (Gibson, 1966; 1979; Treffner & Turvey, 1993). According to this view, the brain is treated as a special purpose device, capable of temporarily adapting its function to specific perception-action situations (Kelso & deGuzman, 1988). In perception, the nervous system may adapt endogenous modes of temporal organization to external rhythmic patterns, controlling attention and memory (Jones, 1976). Other connectionists have noted the fundamental consonance of such dynamical systems approaches with modern connectionist cognitive modeling (e.g. van Gelder & Port, forthcoming). This dissertation is offered in an attempt to bring the two closer together to overcome the limitations of current connectionist models. I have found music perception to be a fertile testing ground for this approach. The current proposal attempts to explain the mechanisms underlying temporal adaptation in the human response to musical rhythms. I believe that this approach will lead to more robust and parsimonious theories of temporal sequence processing in natural and artificial neural networks.

LIST OF REFERENCES

Abraham, R. H., & Shaw, C. D. (1992). *Dynamics: The geometry of behavior (2nd ed.)*. Redwood City, CA: Addison-Wesley.

Allen, P. E., & Dannenberg, R. B. (1989). *Tracking musical beats in real time*. In Proceedings of the 1990 International Computer Music Conference (pp. 140-143). Computer Music Association.

Anderson, S. & Port, R. (1990). A neural model of auditory pattern recognition. Technical Report 1, Institute for the Study of Human Capabilities, Indiana University, Bloomington, IN.

Angeline, P. J., & Pollack, J. B. (1990). Hierarchical RAAMs. Technical Report 91-PA-HRAAMS, Laboratory for Artificial Intelligence Research, The Ohio State University, Columbus, OH.

Apel, W. (1972). *Harvard dictionary of music (2nd ed.)*. Cambridge, MA: Belknap Press of Harvard University Press.

Beek, P. J., Peper, C. E., & van Wieringen, P. C. W. (1992). Frequency locking, frequency modulation, and bifurcations in dynamic movement systems. In G.E. Stelmach and J. Requin (Eds.) *Tutorials in motor behavior II*. Elsevier Science Publishers B. V.

Bengtsson, I. & Gabrielsson, A. (1983). Analysis and synthesis of musical rhythm. In J. Sundberg (Ed.), *Studies of music performance* (pp. 27-60). Stockholm: Royal Swedish Academy of Music.

Bharucha, J.J, & Todd, P.M. (1991). Modeling the perception of tonal structure with neural nets. In P. M. Todd and D. G. Loy (Eds) *Music and Connectionism* (pp. 129-137). Cambridge: MIT Press.

Bodenhausen, U., & Waibel, A. (1991). *The Tempo 2 algorithm: Adjusting time delays by supervised learning*. In R. P. Lippmann, J. Moody, & D. S. Touretsky (Eds) Advances in Neural Information Processing Systems 3. San Mateo, CA: Morgan Kaufman.

Brown, J. C. (1992). Determination musical meter using the method of autocorrelation. *Journal of the Acoustical Society of America*, *91*, 2374-2375.

Brown, J. C. (1993). Determination of the meter of musical scores by autocorrelation. *Journal of the Acoustical Society of America*, *94*, 1953-1957.

Brown, J. C., & Puckette, M. S. (1989). Calculation of a narrowed auto-correlation function. *Journal of the Acoustical Society of America*, *85*, 1595-1601.

Burr, D. & Miyata, Y. (1993). Hierarchical recurrent networks for learning musical structure. In C. Kamm, G. Kuhn, B. Yoon, S. Y. King, & R. Chellappa (Eds.) *Neural Networks for Signal Processing III*. Picataway, NJ: IEEE.

Butler, D. (1992). *The musician's guide to perception and cognition*. NY: Schirmer Books.

Carpenter, G. A., & Grossberg, S. (1983). A neural theory of circadian rhythms: The gated pacemaker. *Biological Cybernetics*, *48*, 35-59.

Carpenter, G. A., & Grossberg, S. (1990). ART 2: Self-organization of stable category recognition codes for analog input patterns. *Applied Optics*, *26*, 4919-4930.

Carpenter, G. A., & Grossberg, S. (1990). ART 3: Hierarchical search using chemical transmitters in self-organizing pattern recognition architectures. *Neural Networks*, *3*, 129-152.

Chalmers, D. (1990). Syntactic transformations on distributed representations. *Connection Science*, *2*, 53-62.

Chrisman, L. (1991). Learning recursive distributed representations for holistic computation. *Connection Science*, *3*, 345-366.

Clarke, E. F. (1985). Structure and expression in rhythmic performance. In P. Howell, I. Cross, and R. West (Eds) *Musical Structure and Cognition*. London: Academic Press.

Clarke, E. F. (1989). The perception of expressive timing in music. *Psychological Research*, *51*, 2-9.

Clarke, E. F. (1993). Imitating and evaluating real and transformed musical performances. *Music Perception*, *10*, 317-341.

Cleeremens, A., Servan-Schreiber, D., & McClelland, J. L. (1989). Finite state Automata and simple recurrent networks. *Neural Computation*, *1*, 372-381.

Cohen, M. A., & Grossberg, S. (1987). Masking fields: A massively parallel neural architecture for learning, recognizing, and predicting multiple groupings of patterned data. *Applied Optics*, *26*, 1866-1891.

Cooper, G., & Meyer, L. B. (1960). *The rhythmic structure of music*. Chicago: University of Chicago Press.

Cottrell, G. W., & Tsung, F. S. (1991). Learning simple arithmetic procedures, In J. A. Barnden and J. B. Pollack (Eds.), *High Level Connectionist Models*. Norwood, NJ: Ablex Publishing.

Cottrell, G. W., Nguyen, M., & Tsung, F. (1993). *Tau Net: The way to do is to be*. In Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society. Hillsdale, NJ:Erlbaum Press.

Cummins, F., Port, R., McAuley, J. D., & Anderson, S. (1993). A neural model of auditory pattern recognition. Technical Report 93, Cognitive Science Program, Indiana University, Bloomington, IN.

Cutler, A., & Mehler, J. (1993). The periodicity bias. *Journal of phonetics*, *21*, 103-108.

Dannenberg, R. B. (1984). *An on-line algorithm for real-time accompaniment*. In Proceedings of the 1984 International Computer Music Conference. Computer Music Association.

Dannenberg, R. B., & Mont-Reynaud, B. (1987). *Following an improvisation in real time*. In Proceedings of the 1987 International Computer Music Conference. Computer Music Association.

de Vries, B., & Principe, J. C. (1992). The gamma model – A new neural net model for temporal processing. *Neural Networks*, *5*, 565-576.

Desain, P., & Honing, H. (1991). The quantization of musical time: A connectionist approach. In P. M. Todd and D. G. Loy (Eds) *Music and Connectionism* (pp. 150-160). Cambridge: MIT Press.

Deutsch, D. (1980). The processing of structured and unstructured tonal sequences. *Perception and Psychophysics*, *28*, 381-389.

Deutsch, D., & Feroe, F. (1981). Internal representation of pitch sequence in tonal music. *Psychological Review*, *88*, 503-522.

Dowling, W. J., & Harwood, D. L. (1986). *Music cognition*. San Diego: Academic Press.

Drake, C., & Botte, M. (1993). Tempo sensitivity in auditory sequences: Evidence for a multiple-look model. *Perception and Psychophysics*, *54*, 277-286.

Drake, C., & Palmer, C. (1993). Accent structures in music performance. *Music Perception*, *10*, 343-378.

Eckhorn, R., Bauer, R., Jordan, W., Brosch, M., Kruse, W., Munk, M. & Reitboeck, H. J. (1988). Coherent oscillations: A mechanism of feature linking in the visual cortex? *Biological Cybernetics*, *60*, 121-130.

Eckhorn, R., Reitboeck, H. J., Arndt, M. & Dicke, P. (1989). *Feature linking via stimulus evoked oscillations: Experimental results from cat visual cortex and functional implications from a network model*. In Proceedings of the International Joint Conference on Neural Networks. Washington.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179-211.

Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grmmatical structure. *Machine Learning*, *7*, 195-224.

Elman, J. L., & Zipser, D. (1988). Learning the hidden structure of speech. *Journal of the Acoustical Society of America*, *83*, 1615-1625..

Essens, P. J., & Povel, D. (1985). Metrical and nonmetrical representation of temporal patterns. *Perception and Psychophysics*, *37*, 1-7.

Estes, W.K. (1972). An associative basis for coding and organization in memory. In A.W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp.161-190). NY: Halsted.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*, 3-71.

Fowler, C. A. (1983). Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in sequences of monosyllabic stress feet. *Journal of Experimental Psychology: General*, *112*, 386-412.

Fowler, C. A. (1990). Sound producing sources as objects of perception: Rate normalization and nonspeech perception. *Journal of the Acoustical Society of America*, *88*, 1236-1249.

Fraisse, P. (1956). Les structures rhythmiques. Louvain: Publication Universitaires de Louvain.

Fraisse, P. (1982). Rhythm tempo. In D. Deutsch (Ed) *The Psychology of Music* (pp. 149-180). NY: Academic Press.

Garner, W. R., & Gottwald, R. L., (1968). The perception and learning of temporal patterns. *Quarterly Journal of Experimental Psychology*, *20*, 97-109.

Getty, D. J. (1975). Discrimination of short temporal intervals: A comparison of two models. *Perception & Psychophysics*, *18*, 1-8.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.

Giles, C. L., Sun, G. Z., Chen, H. H., Lee, Y. C., & Chen, D. (1990). *Higher order recurrent networks and grammatical inference*. In J. E. Moody, S. J. Hanson, & R. P. Lippmann (Eds.), Advances in Neural Information Processing Systems 2. San Mateo, CA: Morgan Kaufman.

Gjerdingen, R. O. (1989). Meter as a mode of attending: A network simulation of attentional rhythmicity in music. *Integral*, *3*, 64-92.

Gjerdingen, R. O. (1990). Catergorization of musical patterns by self-organizing neuronlike network. *Music Perception*, *8*, 339-370.

Gjerdingen, R. O. (1991). Using connectionist models to explore complex musical patterns. In P. M. Todd and D. G. Loy (Eds) *Music and Connectionism* (pp. 138-149). Cambridge: MIT Press.

Glass, L., & Mackey, M. C. (1988). *From clocks to chaos: The rhythms of life*. Princeton, NJ: Princeton University Press.

Gray, C. M., Konig, P., Engel, A. K., & Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*, *338*, 334-337.

Grossberg, S. (1976). Adaptive pattern classification and universal recording II: Feedback, expectation, olfaction, and illusions. *Biological Cybernetics*, *23*, 187-202.

Halpern, A. R., & Darwin, C. (1982). Duration discrimination in a series of rhythmic events. *Perception & Psychophysics*, *31*, 86-89.

Handel, S. (1989). *Listening*. Cambridge, MA: MIT Press.

Jackendoff, R. (1992) Musical processing and musical affect. In M. R. Jones and S. Holleran (Eds.) *Cognitive bases of musical communication*. Washington: American Psychological Association.

Johnson-Laird, P. N. (1991). Jazz improvisation: a theory at the computational level. In P. Howell, R. West, and I. Cross (Eds.), *Representing musical structure*. London: Academic Press.

Jones, M. R. (1974). Cognitive representations of serial patterns. In B.H. Kantowitz (Ed.), *Human information processing*. NY: John Wiley & Sons.

Jones, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review*, *83*, 323-335.

Jones, M. R. (1981). A tutorial on some issues and methods in serial pattern research. *Perception & Psychophysics*, *30*, (5), 492-504.

Jones, M. R. (1981). Music as a stimulus for psychological motion: Part I: Some determinants of expectancies. *Psychomusicology*, *1*, 34-51.

Jones, M. R. (1987). Dynamic pattern structure in music: Recent theory and research. *Perception & Psychophysics*, *41*, 621-634.

Jones, M. R., & Boltz, M. (1989). Dynamic Attending and Responses to Time. *Psychological Review*, *96*, 459-491.

Jones, M. R., Boltz, M., & Kidd, G. (1982). Controlled attending as a function of melodic and temporal context. *Journal of Experimental Psychology: Human Perception & Performance*, *7*, 211-218.

Jones, M. R., Kidd, G., Wetzel, R. (1981). Evidence for rhythmic attention. *Journal of Experimental Psychology: Human Perception & Performance*, *7*, 1059-1073.

Jones, T., Laine, P., Tiits, K., Torkkola, K. (1981). A nonheuristic automatic composing method. In P. M. Todd and D. G. Loy (Eds) *Music and Connectionism* (pp. 229-242). Cambridge: MIT Press.

Jordan, M. (1986). Serial order. Technical Report 8604, Institute for Cognitive Science, University of California at San Diego, La Jolla, CA.

Kelso, J. A. S., & deGuzman, G. C. (1988). Order in time: How the cooperation between the hands informs the design of the brain. In H. Haken (Ed.). *Natural and Synergetic Computers*. Berlin: Springer-Verlag.

Knopoff, L., & Hutchinson, W. (1978). An index of melodic activity. *Interface*, *7*, 205-229.

Kohonen, T. (1978). *Self-organization and associative memory.* Berlin: Springer-Verlag.

Kolen, J. F. (1994). Exploring the computational capabilitiesof recurrent neural networks. Unpublished Ph.D. Dissertation, The Ohio State University, Columbus, OH.

Krumhansl, C. L., Bharucha, J. J., & Castellano, M. A. (1982). Key distance effects on perceived harmonic structure in music. *Perception & Psychophysics*, *32*, 96-108.

Krumhansl, C.L. (1979). The psychological representation of musical pitch in a tonal context. *Cognitive Psychology*, *11*, 346-384.

Lang, K., Waibel, A. H., & Hinton, G. E. (1990). A time-delay neural network architecture for isolated word recognition. *Neural Networks*, *3*, 23-44.

Lapedes & Farber, (1987). ????.

Large, E. W. (1992). *Judgements of similarity for musical sequences*. Unpublished manuscript, Center for Cognitive Science. The Ohio State University, Columbus, OH.

Large, E. W., & Kolen, J. F. (1993). *A dynamical model of the perception of metrical structure*. Presented at Society for Music Perception and Cognition. Philadelphia, June.

Large, E. W., & Kolen, J. F. (in press). Resonance and the perception of musical meter. *Connection Science*.

Large, E. W., Palmer, C., & Pollack, J. B. (1991). *A connectionist model of intermediate representations for musical structure*. In Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society (pp.412 - 417). Hillsdale, N.J.:Erlbaum Press.

Large, E. W., Palmer, C., & Pollack, J. B. (in press). Reduced memory representations for music. *Cognitive Science*.

Lashley, K. (1951). The problem of serial order in behavior. In Jeffress (Ed.). *Cerebral mechanisms in behavior.* NY: Wiley.

Lehistse, I. (1977). Isochrony reconsidered. *Journal of phonetics*, *5*, 253-263.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge: MIT Press.

Liberman, M. (1975). *The intonational system of English*. Unpublished Ph.D. Dissertation, MIT.

Longuet-Higgins, H. C. (1987). *Mental Processes*. Cambridge: MIT Press.

Longuet-Higgins, H. C., & Lee, C. S. (1982). The perception of musical rhythms. *Proceeding of the Royal Society of London B*, *207*, 187-217.

Marr, D. (1982). *Vision*. New York: W. H. Freeman.

Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Perception*, *11*, 115-128.

McAuley, J. D. (1993). *Learning to perceive and produce rhythmic patterns in an artificial neural network*. Technical Report, Department of Computer Science, Indiana University.

McAuley, J. D. (1994). *Finding metrical structure in time*. In M. C. Mozer, P. Smolensky, D. S. Touretsky, J. L. Elman & A. S. Weigend (Eds) Proceedings of the 1993 Connectionist Models Summer School. Hillsdale, NJ: Erlbaum Associates.

McGraw, G., Montante, R., & Chalmers, D. (1991). *Rap-master network: Exploring temporal pattern recognition with recurrent networks*. Technical Report No. 336. Computer Science Department, Indiana University.

Meyer, L. (1956). *Emotion and meaning in music*. Chicago: University of Chicago Press.

Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81-97.

Mozer, M. C. (1989). A focused back-propgation algorithm fot temporal pattern recognition. *Complex Systems*, *3*, 349-381.

Mozer, M. C. (1991). Connectionist music composition based on melodic, stylistic and psychophysical constraints. In P. M. Todd and D. G. Loy (Eds) *Music and Connectionism* (pp. 195-211). Cambridge: MIT Press.

Mozer, M. C. (1992). *Induction of multiscale temporal structure*.In R. P. Lippmann, J. Moody, & D. S. Touretsky (Eds.), Advances in Neural Information Processing Systems 4 (pp. 275-282). San Mateo, CA: Morgan Kaufman.

Mozer, M. C. (1993). Neural net architectures for temporal sequence processing. In A. Weigand N. Gershenfeld (Eds.), *Predicting the future and understanding the past* (pp. 243-264). Reading, MA: Addison-Wesley.

Mozer, M. C. (in press). Neural net music composition by prediction: Exploring the benefits of psychoacoustic constraints and multiscale processing. *Connection Science*.

Narmour, E.(1990). *The analysis and cognition of basic melodic structures: The implication-realization model*. Chicago: University of Chicago Press.

Nigrin, A. L. (1990). Stable learning of temporal patterns with an adaptive resonance circuit. Unpublished doctoral dissertation, Duke University.

Oppenheim, A. V., & Schafer, R. W. (1975). *Digital Signal Processing*. Englewood Cliffs, NJ: Prentice Hall.

Page, M. P. A. (1993). Modelling aspects of music perception using self-organizing neural networks. Unpublished doctoral dissertation, University of Cardiff.

Palmer, C. (1988). Timing in skilled music performance. Unpublished doctoral dissertation, Cornell University, Ithaca, NY.

Palmer, C., & Krumhansl, C. L. (1987a). Pitch and temporal contributions to musical phrase perception: Effects of harmony, performance timing, and familiarity. *Perception and Psychophysics*, *41*, 505-518.

Palmer, C., & Krumhansl, C. L. (1987b). Independent temporal and pitch structures in determination of musical phrases. *Journal of Experimental Psychology: Human Perception & Performance*, *13*, 116-126.

Palmer, C., & Krumhansl, C.L. (1990). Mental representations of musical meter. J*ournal of Experimental Psychology: Human Perception & Performance*, *16*, 728-741.

Pollack, J. B. (1988). *Recursive auto-associative memory: Devising compositional distributed representations*. In Proceedings of the Tenth Annual Conference of the Cognitive Science Society (pp. 33-39). Hillsdale, N.J.:Erlbaum Press.

Pollack, J. B.(1990). Recursive distributed representations. *Artificial Intelligence*, *46*, 77-105.

Pollack, J. B.(1991). The induction of dynamical recognizers. *Machine Learning*, *7*, 227-252.

Povel, D., & Essens, P. J. (1985). Perception of temporal patterns. *Music Perception*, *2*, 411-440.

Povel, D., & Okkerman, H. (1981). Accents in equitone sequences. *Perception& Psychophysics*, *7*, 565-572.

Pressing, J. (1988). Improvisation: methods and models. In J. Sloboda (Ed.) *Generative processes in music: the psychology of performance, improvisation, and composition* (pp.129-178). NY: Oxford University Press.

Principe, J. C., de Vries, B., De Oliveira, P. G. (1993). The gamma filter – A new class of adaptive IIR filters with restricted feedback. *IEEE Transactions on Signal Processing*, *41*, 649-656.

Restle, F. (1970). Theory of serial pattern learning: Structural trees. *Psychological Review*, *77*, 481-495.

Rosenthal, D. (1992). Emulation of human rhythm perception. *Computer Music Journal*, 16, 64-76.Rumelhart, D. E., & Zipser, D. (1985). Feature discovery by competitive learning, In D. E. Rumelhart & J. L. McClelland (Eds.) *Parallel distributed processing* . Cambridge: MIT Press.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation, In D. E. Rumelhart & J. L. McClelland (Eds.) *Parallel distributed processing* (pp.318-362). Cambridge: MIT Press.

Scarborough, D. L., Miller, P., & Jones, J. A. (1992). On the perception of meter. In M. Balaban, K. Ebcioglu, & O. Laske (Eds) *Understanding Music with AI: Perspectives in Music Cognition* (pp. 427-447). Cambridge: MIT Press.

Schenker, H. (1979). *Free composition* (E. Oster, Trans.). NY: Longman.

Schmidt, R. C., Beek, P. J., Treffner, P. J., & Turvey, M. T. (1991). Dynamical substructure of coordinated rhythmic movements. *Journal of Experimental Psychology: Human Perception & Performance*, *17*, 635 - 651.

Schmidt, R. C., Shaw, B. K., & Turvey, M. T. (1993). Coupling dynamics in interlimb coordination. *Journal of Experimental Psychology: Human Perception & Performance*, *19*, 397 - 415.

Schroeder, M. (1991). *Fractals, Chaos, Power Laws*. New York: W. H. Freeman and Company.

Schulze, H. H. (1989). The perception of temporal deviations in isochronic patterns. *Perception & Psychophysics*, *45*, 291-296.

Selkirk, E. (1978). On prosodic structure and its relation to syntactic structure. Unpulished Ph.D. Dissertation, University of Massachusetts, Amherst.

Serafine, M. L., Glassman, N., & Overbeeke, C. (1989). The cognitive reality of hierarchic structure in music. *Music Perception*, *6*, 347-430.

Shaffer, L. H. (1981). Performances of Chopin, Bach, and Bartok: Studies in motor programming. *Cognitive Psychology*, *13*, 326-376.

Shaffer, L. H. (1981). Performances of Chopin, Bach, and Bartok: Studies in motor programming. *Cognitive Psychology*, *13*, 326-376.

Shaffer, L. H., Clarke, E., & Todd, N. P. M. (1985). Metre and rhythm in piano playing. *Cognition*, *20*, 61-77.

Simon, H. A., & Kotovsky, K. (1963). Human acquisition of concepts for sequential patterns. *Psychological Review*, *79*, 369-382.

Simon, H. A., & Sumner, R. K. (1968). Pattern in music. In B. Kleinmuntz (Ed.). *Formal Representation of Human Thought* . New York: Wiley.

Sloboda, J. A. (1983). The communication of musical metre in piano performance. *Quarterly Journal of Experimental Psychology*, *35*, 377-396.

Sloboda, J. A. (1985). *The musical mind*. Oxford: Oxford University Press.

Sperdutti, D. (1993). Representing symbolic data structures using neural networks. Unpublished doctoral dissertation, University of Pisa, Pisa, Italy.

Steedman, M. (1982). A generative grammar for jazz chord sequences. *Music Perception*, *2*, 52-77.

Stevens, C. & Wiles, J. (1994). *Tonal music as a componential code: Leraning temporal relationships between and within pitch and timing components*. In R. P. Lippmann, J. Moody, & D. S. Touretsky (Eds.), Advances in Neural Information Processing Systems 6 (pp. 1085-1092). San Mateo, CA: Morgan Kaufman.

Tank, D. W., & Hopfield, J. J. (1987). Neural computation by concentrating information in time. *Proceedings of the National Academy of Sciences*, *81*, 1896-1900.

Todd, N. P. M. (1985). A model of expressive timing in tonal music. *Music Perception*, *3*, 33-59.

Todd, N. P. M. (1994). The auditory primal sketch: A multi-scale model of rhythmic grouping. *Journal of New Music Research*, *23*, 25-69.

Todd, P. M. (1991). A connectionist approach to algorithmic composition. In P. M. Todd and D. G. Loy (Eds) *Music and Connectionism* (pp. 173-194). Cambridge: MIT Press.

Torras, C. (1985). *Temporal pattern learning in neural models*. Berlin: Springer-Verlag.

Treffner, P. J., & Turvey, M. T. (1993). Resonance constraints on rhythmic movement. *Journal of Experimental Psychology: Human Perception & Performance*, *19*, 1221-1237.

Unnikrishnan, K. P., Hopfield, J. J., & Tank, D. W. (1991). Connected-digit speaker-dependent speech recognition using a neural network with time-delayed connections. *IEEE Transactions on Signal Processing*, *39*, 698-713.

van Gelder, T., & Port, R. (forthcoming). Introduction. In T. van Gelder & R. Port (Eds) *Mind as Motion*. Cambridge: MIT Press.

Vercoe, B., & Puckette, M. (1985). *Synthetic rehearsal: Training the synthetic performer.* In Proceedings of the 1985 International Computer Music Conference (pp. 275-278). Computer Music Association.

Vitz, P.C., & Todd, T.C. (1969). A coded element model of the perceptual processing of sequential stimuli. *Psychological Review*, *76*, 433-449.

von der Malsberg, C., & Schneider, W. (1986). A neural cocktail-party processor. *Biological Cybernetics*, *54*, 29-40.

Waibel, A., Toshiyuki, H., Hinton, G. Shikano, K., Lang, K., D. (1989). Phoneme recognition using time-delay neural networks. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *23*, 993-1009.

Wang, D. L. (1993). Pattern recognition: Neural networks in perspective. *IEEE Expert*, *8*, 52-60.

Wang, D. L. (in press). Emergent synchrony in locally coupled neural oscillators. *IEEE Transactions on Neural Networks.*

Wang, D. L. (in press). Temporal pattern processing in neural networks. In M. A. Arbib (Eds.) *Handbook of Brain Theory and Neural Networks*. Cambridge: MIT Press.

Wang, D. L., & Arbib, M. A. (1993). Timing and chunking in processing temporal order. *IEEE Transactions on Systems, Man, and Cybernetics*, *23*, 993-1009.

Wang, D. L., & Arbib, M. A.. (1990). Complex temporal sequence learning based on short-term memory. *Proceedings of IEEE*, *78*, 1536-1543.

Winfree, A. T. (1980). *The geometry of biological time*. New York: Springer-Verlag.

Woodrow, H. (1909). A quantitative study of rhythm. *Archives of Psychology*, *14*, 1-66.

Woodrow, H. (1951). Time perception. In S. S. Stevens (Ed.) *Handbook of Experimental Psychology* (pp.1224-1236) New York: Wiley.

Yee, W., Holleran, S., & Jones, M. R. (in press). Sensitivity to event timing in regular and irregular sequences: Influences of musical skill. *Journal of Experimental Psychology: Human Perception & Performance*.

Yeston, M. (1976). *The stratification of musical rhythm*. New Haven: Yale University Press.

Zanzone, P. G., & Kelso, J. A. S. (1992). Evolution of behavioral attractors with learning: Nonequilibrium phase transitions. *Journal of Experimental Psychology: Human Perception & Performance*, *18*, 403-421.