

An Agency for the Perception of Musical Beats
or
If I Only Had a Foot...

by

Joseph Taisup Chung

Submitted to the Department of
Electrical Engineering and Computer Science
in Partial Fulfillment of the Requirements for the Degrees of

Bachelor of Science
and
Master of Science in Computer Science

at the
Massachusetts Institute of Technology
June, 1989

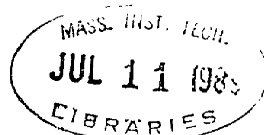
© Massachusetts Institute of Technology, 1989. All Rights Reserved

Signature of Author _____
Department of Electrical Engineering and Computer Science
May 23, 1989

Certified by _____
Tod Machover
Associate Professor of Music and Media
Thesis Supervisor

Certified by _____
Marvin Minsky
Donner Professor of Science
Thesis Supervisor

Accepted by _____
Arthur C. Smith
Chairman, Committee on Graduate Students



ARCHIVES

**An Agency for the Perception of Musical Beats
or
If I Only Had a Foot...**

by

Joseph Taisup Chung

Submitted to the Department of
Electrical Engineering and Computer Science
in Partial Fulfillment of the Requirements for the Degrees of
Bachelor of Science and Master of Science in Computer Science
at the Massachusetts Institute of Technology
June, 1989

Abstract

In many kinds of music, listeners readily perceive a periodic beat or pulse which is strongly correlated with the onsets of notes and the pattern of musical accents. This beat is often evidenced by the ability of listeners to tap their feet, and plays an important role in the perception of musical structure. This thesis presents a real time model of human beat perception as an *agency* based on Marvin Minsky's theories of the mind. This model is implemented as a computer program which accepts real performances of monophonic music (in the form of MIDI) and discovers the beats.

Thesis Supervisors: Tod Machover
 Associate Professor of Music and Media

 Marvin Minsky
 Donner Professor of Science

Acknowledgments

I would first like to thank my advisor, friend, and mentor, Tod Machover, who has taught me so many things that I can hardly begin to name them. For the encouragement and support, the advice in so many matters, and of course, the Music, thank you.

I would also like to thank Marvin Minsky for co-supervising this thesis and providing the basic ideas, and Nicholas Negroponte for creating the environment here at the Media Laboratory where this work could take place. Further thanks to Dave Rosenthal, Alan Ruttenberg, and Robert Rowe for all the patient help and useful discussions, and to Michael Hawley for the performances of the musical examples.

I'd also like to say how much I have appreciated the presence of my Media Lab friends who made this place bearable at 4am, especially, Kael, Swimfast, ML CoolJ, jh, and Straz. And Tom.

Lastly, I'd like to thank Faith for putting up with me since I've been in work-mode, and for getting me into work-mode in the first place.

This thesis is dedicated to my Mother and Father.

Table of Contents

Abstract.....	2
Acknowledgments.....	3
Table of Contents	4
Introduction.....	6
The Perception of Musical Beats: An Overview.....	8
Beats and Meter.....	8
Beat Perception of Perfectly Mechanical Performances.....	12
Beat Perception of Real Performances.....	15
Ambiguity.....	18
Rhythmic Ambiguity vs. Metric Ambiguity.....	20
Complexity	21
Rhythm Perception vs. Music Perception.....	26
Metrical Preference Rules	33
The Beat Perception Process	36
Expectation	37
Disambiguation.....	40
Beat Perception of Perfectly Mechanical and Real Performances Revisited.....	42
Summary.....	44
Previous Work in Beat Perception Machines	45
Kinds of Theories.....	45
Constraints on Perceptual Theories.....	46
Type of Input.....	48
Review of Previous Work.....	48
Longuet-Higgins and Lee (1984).....	49
Steedman (1977).....	49
Dannenberg and Mont-Reynaud (1978).....	50
Chowning et. al. (1984) and Schloss (1985)	51
C.S. Lee (1985).....	52
An Agency for the Perception of Musical Beats.....	56
Society of (Musical) Mind	56
Agents and Agencies.....	56
What Are Beats and Meter Good For?.....	58
A Beat Perception Agency.....	60
Musical Input.....	61
The Note Importance Agency.....	61
The Metrical Hierarchy Agency.....	63
Beat-level Agents.....	63
Profusion of Beat-levels	66
Scoring Beat-levels	67
Reconverging Beat-Levels	68
Meter Agencies.....	71

Superior, Subordinate and Top-Level Meter Agencies.....	72
Scoring Meter Agencies.....	74
Creating Meter Agencies.....	75
Creating Simple Beat-Levels.....	77
Inducing Higher Levels.....	77
Attaching Higher Levels.....	80
Managing Meter Agencies.....	81
Implementation.....	82
Input.....	83
Note Importance.....	83
Objects and Agencies.....	83
Discussion.....	83
Rhythmic and Metric Interpretation.....	83
Ambiguity.....	85
A "Society of Mind" Model.....	86
Problems.....	87
Conclusions.....	90
Future Research Directions.....	91
Note Importance Agency.....	91
Non-Monophonic Music.....	91
Live Beat Tracking System.....	92
Possible Applications.....	92
Automatic Transcription.....	92
Hyperinstruments.....	92
Appendices.....	94
Hyperlisp.....	94
Bibliography.....	97

Introduction

One often thinks of music as staves and notes on a page, forgetting that the real music is an entirely perceptual phenomenon. The things we call notes, melodies, and rhythms are abstractions for certain cognitive processes which exist only in our minds. To study music, therefore, is to study human thought. While the advent of techniques for manipulating representations of music on a computer has provided music researchers with powerful tools, a profusion of “musically intelligent” systems is still forthcoming. I believe that the difficulty encountered in programming computers to perform beat tracking, automatic transcription, “name the composer,” and so forth, can be attributed to the lack of a useful paradigm of human performance of these processes.

A computer which is analyzing a musical performance represented at the “note-level,” such as in MIDI (see MIDI Specification 1.0, 1985), can easily find the average time interval between note onsets (the *inter-onset* interval) and the exact number of times each note is played. Yet, the computer cannot tell us many things that are easily evident to untrained listeners, such as which notes are on the beat, or what the melody is (if it exists). There is an identifiable difference between these two kinds of musical problems: the easy problems concern the properties of the representation (in this case MIDI), whereas the hard problems concern the properties of human perceptual processes. The hard problems are questions about mental structures that are inferred, but separate from the concrete representations.

The goal of this thesis is to solve a specific musical problem, namely the construction of a machine which can “tap its foot” to the beat, but the approach taken is to model the mental structures of beat perception within the *Society of Mind* framework established by Marvin Minsky (1981 & 1986). The premise is that beat perception as well as other “hard” musical problems can be solved if the computer representation of music correctly models the human dynamic processes of listening.

The Perception of Musical Beats: An Overview

Beats and Meter

The work in this thesis applies to music which is considered to be *metrical*, that is, music to which there is both a perceivable, periodic pulse, and a perceivable, periodic higher level organization to the pulse derived from a pattern of recurring musical accents (Cooper & Meyer 1960). This higher level organization must form a new periodic pulse whose beats coincide with the beats of the lower level, but whose period is an integral multiple of the lower level (usually two or three). In general, metrical music tends to have several perceivable beat levels, forming a *metrical hierarchy*.

A graphical depiction of a metrical hierarchy can be found in Lerdahl and Jackendoff's *A Generative Theory of Tonal Music* (Lerdahl & Jackendoff 1983). Each beat at a particular level is shown with a dot. A beat that is felt to be accented, or "strong" at some level becomes a beat at the next higher level. An example of the hierarchy of 4/4 meter reproduced below:¹

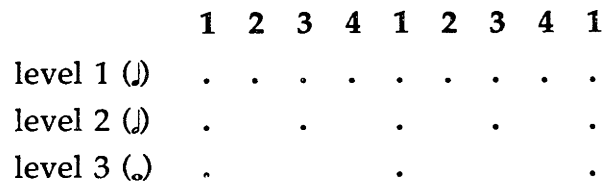


Figure 1.

¹ Lerdahl and Jackendoff (1985), p. 19.

In this example, every other beat beginning with the first is felt to be “strong” at level one and thus becomes a beat at level two. At level two, again every other beat is “strong”, and becomes a beat at level three.

According to Lerdahl and Jackendoff:

The listener tends to focus primarily on one (or two) intermediate level(s) [of the beat hierarchy] in which the beats pass by at a moderate rate. This is the level at which the conductor waves his baton, the listener taps his foot, and the dancer completes a shift in weight.²

At first glance, beat perception may seem to be an independent sub-problem of meter perception. It may seem that we first identify which notes are beats and then abstract a metrical hierarchy afterwards. Lerdahl and Jackendoff, for example, claim that the metrical hierarchy is derived entirely from the regularity of phenomenal accents which occur “at the musical surface”:

Phenomenal accent functions as a perceptual input to metrical accent – that is, the moments of musical stress in the raw signal serve as “cues” from which the listener attempts to extrapolate a regular pattern of metrical accents.³

Lerdahl and Jackendoff define *phenomenal accents*, as “any event at the musical surface that gives emphasis or stress to a moment in the musical flow.”⁴ They include in this category, “sforzandi, sudden changes in dynamics

² Lerdahl and Jackendoff (1985), p. 17.

³ Ibid.

⁴ Ibid.

or timbre, long notes, leaps to relatively high or low notes, harmonic changes and so forth.”⁵

In actuality, Lerdahl and Jackendoff back away significantly from this claim in their more detailed discussion of metrical preference rules, allowing for the influence of deeper structural information. For example, their first metric preference rule states “Where two or more groups or parts of groups can be construed as parallel, they preferably receive parallel metrical structure.”⁶ Certainly comparisons of parallelism involve consideration of many aspects of musical structure that are not “at the musical surface.” These preference rules are discussed at length in a later section.

The important point is that the perception of beats and meter is neither a simple nor obvious process which need only examine a narrow dimension of easily accessible information. Cooper and Meyer (1960) support this claim in their discussion of accent, which differs considerably from the concept of Lerdahl and Jackendoff’s phenomenal accent. Cooper and Meyer take the accentedness of a note essentially as a given, irrespective of its relative length or loudness. Thus, for metrical music, accent becomes the property of a note coinciding with a beat, much like Lerdahl and Jackendoff’s notion of metrical accent. Cooper and Meyer state:

... One cannot at present state unequivocally what makes one tone seem accented and another not. For while such factors as duration, intensity, melodic contour, regularity, and so forth obviously play a part in creating an impression of accent, none of them appears to be an

⁵ Lerdahl and Jackendoff (1985), p. 17.

⁶ *Ibid.*, p. 75.

invariable and necessary concomitant of accent. Accents may occur on short notes as well as long, on soft notes as well as loud, on lower notes as well as higher ones, and irregularly as well as regularly.⁷

This thesis shall argue that, in general, the process of correctly identifying beats, rhythmic values and metrical hierarchy in a real performance of music is a complex, circularly dependent process which is influenced by many factors, especially the various musical expectations that are established in the course of the listening experience. I shall show that while information at the musical surface (such as phenomenal accents as defined above) provides important metrical cues, these cues can only indicate a vast number of ambiguous metric interpretations. Deeper structural processes of perception are required to disambiguate the indicated interpretations.

In order to explore the various aspects of this problem, I shall begin by considering a special sub-class of musical performances, which I shall call *perfectly mechanical*. A perfectly mechanical performance can be defined as a performance in which the timing of all notes is precisely as notated, all notes are played exactly at the same loudness and with the same articulation, and the tempo is perfectly maintained. In such a case, there are no timing ambiguities whatsoever: each note value (quarter note, sixteenth note, eighth note triplet, etc.) has a specific clock time duration. Aside from the pitch of the note, there is no other information available.

⁷ Cooper and Meyer (1960), p. 7.

Beat Perception of Perfectly Mechanical Performances

A perfectly mechanical performance can be thought of as a sequence of unbarred, unmetered notes such as:



Figure 2.

If this sequence were heard in isolation with equal dynamic stress on each note, most listeners would probably hear such a sequence in 3/8 with the downbeat on the quarter notes:



Figure 3.

However, if such a sequence followed music which established a strong 2/4 metrical context, listeners might perceive the same sequence quite differently. Figure 4 illustrates three possible 2/4 interpretations, and figure 5 shows a possible context for this sequence which might give rise to the first of these 2/4 interpretations:

Their principal finding is that listeners tend to choose metric interpretations which result in *regular passages*, essentially passages without syncopation.

Longuet-Higgins and Lee define a regular passage as follows:

1. Every bar, except possibly the first, begins with a sounded note (this ensures that there are no syncopations across bar lines).
2. All the bars are generated by the same standard meter.
3. There are no syncopations within any of the bars.⁸

Longuet-Higgins and Lee relax their third criterion slightly, theorizing that legato phrasing can have the effect of grouping syncopated notes within bars together into a single “virtual note” of equivalent duration. If one replaces the syncopated notes with the virtual note, the resulting passage becomes regular.

Although they do not discuss cases where there is syncopation across bar lines as in figure 4 above, one may assume that interpretations which result in phrases which are as regular as possible are perceptually preferred. Based on this hypothesis, a beat perception machine might attempt to generate all the feasible interpretations which result in passages that are regular or close to regular. However, this process would still result in potentially large numbers of interpretations. Clearly, other criteria besides “regularity” must be used to disambiguate the possibilities.

⁸ Longuet-Higgins & Lee (1984), p. 431.

Before exploring these criteria, it is worthwhile to consider the additional issues associated with real (as opposed to perfectly mechanical) performances, especially with regard to timing.

Beat Perception of Real Performances

One of the most surprising findings in extensive analysis of real musical performances is the extent to which the timing of notes deviates from the perfectly mechanical values. In several studies of performance timing (Gabrielsson 1973; Gabrielsson et al. 1983; Bengtsson & Gabrielsson 1980, 1983), researchers found deviations at times approaching 40% of the notated values for eighths and sixteenth notes. These findings lead Gabrielsson to ask:

Why do these “shocking” deviations from exact frequencies, constant tempo, exact temporal relations, etc. occur in performed music – and why, on the whole, don’t we perceive them as unwanted and distasteful deviations or irregularities?⁹

Researchers have demonstrated that a great deal of timing deviation is systematic, not random. For example, Gabrielsson (1985) claims that “the accompaniment in a Viennese waltz is usually performed with a ‘short first beat’ but a ‘long second beat,’” demonstrating systematic duration deviation at the beat level. Other studies (Bengtsson and Gabrielsson 1983; Shaffer, Clarke, & Todd 1985; Todd 1985) have found systematic lengthening of the durations of notes which end musical phrases (*phrase-final lengthening*).

⁹ Gabrielsson (1985), p. 63.

Shaffer and Todd (1987) found that in two performances by the same pianist of Chopin's Prelude in F Sharp Minor, the timing deviations, though quite significant, correlated extremely well with each other. Shaffer and Todd argue that the high degree of precision in the reproduction of the systematic timing deviations demonstrates "that an expressive form can have a precise mental representation and can be precisely executed."¹⁰

These findings concur with musical intuition: if the timing in a performance is too regular, the result sounds mechanical and unmusical. A good musician intentionally introduces systematic timing deviations from a perfectly mechanical performance which seems to convey important information to the listener (Gabrielsson 1985, 1988; Todd 1985).

In the context of this thesis, where the intent is to build a system which can detect and recover the beats and metrical structure, the timing deviations introduce a further dimension of uncertainty. One cannot, in general, simply map measured time intervals to notated values. The deviations are great enough to easily confuse triplets for eighth notes, dotted quarter notes for half notes, etc.

As an example, consider the following "piano roll" style graphical representation of a real performance of a Bach Chorale (Cantata 140)¹¹:

¹⁰ Shaffer and Todd (1987), p. 142.

¹¹ This example, as well as others in this thesis, was recorded on a Bosendorfer 290 SE recording piano by Michael Hawley, an accomplished, non-professional pianist.

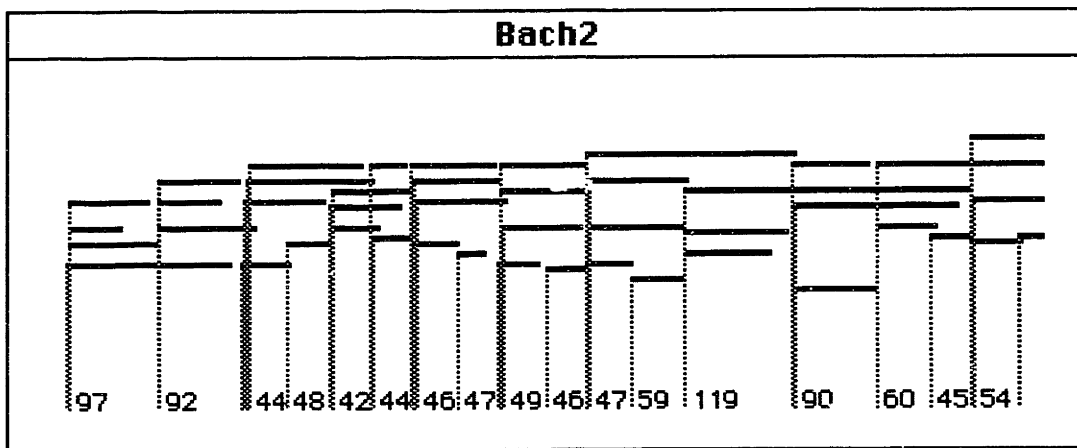


Figure 6.

The numbers at the bottom represent the time between note onsets (the inter-onset interval) measured in 1/100ths of a second (centiseconds). All the notes in the above figure are either quarter-notes or eighth-notes. However, we observe a wide range of time values even in such a short segment of music. Eighth-notes vary in time between 42 and 54 centiseconds, whereas quarter notes range between 90 and 119 centiseconds. By these values, a dotted quarter note may be as small as 132 centiseconds, and a naive attempt at recovering the notated values from the inter-onset intervals might mistake the 119 centisecond quarter note for a dotted quarter.

Because the intended note values are not directly recoverable from the note onset times, timing deviations in real performances create another source of ambiguity. It is a non-trivial task to recover the perfectly mechanical version of a real performance, which limits the direct usefulness of Longuet-Higgins' and Lee's findings. On the other hand, it is a fallacy to assume that recovering the perfectly mechanical performance is a sub-problem of finding the beats

and meter. This subject will be discussed at length in a following section on the beat perception process.

It should be noted that the above musical example is a very simple case, and is not heard as exaggerated or unnatural with regard to timing. Far more complex cases can occur when the music itself is more complex, containing many different note values, dotted figures, triplets, etc., and when the performance is less regular with regard to overall tempo, local tempo, and unintentional timing inaccuracies.

With respect to errors, it is interesting to reflect that as listeners, we are remarkably tolerant. One can listen to a performance of a previously unheard piece which is played badly, even to the point of skipping beats or stopping, and still correctly interpret the beat and meter. Our sense of meter perception is robust enough that we can often identify and recover from gross errors in performance without immediately getting confused or switching interpretations.¹²

Ambiguity

There seems to be an underlying assumption in the literature on rhythm and meter that there is a clear cut division between cases where listeners' perception of a metrical hierarchy is ambiguous, or unambiguous. Additionally, one commonly finds implications that the normal listening case is an unambiguous metrical hierarchy with each listener perceiving the

¹² This reasoning is based only on introspection and intuition. The author was unable to find any applicable studies based on badly performed music.

notated meter. Lerdahl and Jackendoff (1985), for example, cite as an example of unambiguous metrical structure, a “not untypically complex passage: the beginning of Mozart’s G Minor Symphony [Number 40],”¹³ and claim that “the cues in the music from the 8th-note level to the 2-bar level unambiguously support the analysis given.”¹⁴

However, there is evidence that although listeners may not feel any sense of ambiguity, they may not be perceiving the same metrical structure as other listeners. Fraisse (1982) cites the following experiment performed by Peter Vos:

On commercial versions of Bach’s preludes, subjects familiar with classical music but not particularly acquainted with the pieces chosen were asked to tap in synchrony with the beginning of the perceived rhythmic pattern. The subject did not tap in all cases on the first beat of the bar. Let us take the example of a 2/4 bar that lasted 1.75 sec. Forty percent of the subjects tapped in synchrony with the first beat ...; 45% tapped on the second beat ... 10% tapped each beat.¹⁵

Furthermore, even in music that is generally considered to be metrical, there is a wide range in how obvious the metrical structure is intended to be. Such music ranges from music where the meter is intended to be quite explicit, such as popular dance (disco) music, to music where the metrical structure is intended to be somewhat ambiguous. Sloboda (1983) cites the opening of the last movement of Beethoven’s Piano Sonata in G (Opus 14, Number 2) as music that falls in the latter category. In fact, there is always some metrical

¹³ Lerdahl and Jackendoff (1985), p. 22.

¹⁴ Ibid., p. 22.

¹⁵ Fraisse (1982), p. 173.

ambiguity in any piece, at least for the first few bars, before any sort of structure can be inferred.

The relevance of this discussion to the work in this thesis is that while phrases such as “the correct metrical interpretation” may be convenient in describing what a performer and a composer might have intended, listeners’ perception can be highly subjective and variable. Even in simple music, listeners may not perceive the notated metrical structure. While a beat perception machine should be capable of providing its best metrical interpretation, it is equally important that it maintain alternative interpretations and possess some measure by which to compare them. Presumably it is the presence of these alternative interpretations that makes music rhythmically interesting.

Rhythmic Ambiguity vs. Metric Ambiguity

It is important to distinguish between the two different types of ambiguity which we have discussed. The ambiguity of intended note values introduced by timing deviations shall be referred to as *rhythmic ambiguity*, and the ambiguity between possible metrical interpretations inherent in any sequence of notes shall be called *metric ambiguity*. Thus we can speak of two different, but related kinds of interpretations: *rhythmic interpretations* and *metric interpretations*.

A rhythmic interpretation of a passage of music is a theory as to the intended note values, while a metric interpretation is a theory as to the location of the beats and the metrical hierarchy. I shall argue in a later section that obtaining

the correct metric and rhythmic interpretations are part of the same process, and that one cannot perform one without performing the other.

Complexity

If one accepts that human thought takes place entirely in the physical realm of our brains (i.e. that human thought does not embody some sort of supernatural phenomenon), then human thought must be subject to the same laws which govern all computational machines. These laws, usually called *complexity* and *computability* theorems, have to do with what problems are solvable by computational machines and how the complexity of solvable problems can be characterized.¹⁶

The profusion of ambiguous rhythmic and metric interpretations shown in the previous two sections brings up a complexity issue regarding the amount of computation required as a function of the number of musical events. Consider the following figure which graphically shows the note onset times of a musical passage as new notes are heard (case *e* shows the onset times of the entire phrase):

¹⁶ See Lewis & Papadimitiou (1981) for a rigorous examination of these issues.

limit the number of possibilities for this example, we shall only allow quarter-notes, dotted-quarter notes, eighth-notes and sixteenth-notes. We shall, however, allow the amount of deviation to be great enough so that in case *b* we can interpret the two notes as follows:



Figure 9.

When the next note is heard (case *c* above), each of these three interpretations can have two allowable continuations, show below:



Figure 10.

With the onset of the next note (case *d* above), the above interpretations can again be continued in up to two ways:



Figure 11.

Even in such a simple case (the onset times are from the opening bars of Mozart's "Eine kleine Nachtmusik") one can see that since every note tends to have more than one interpretation, the total number of interpretations multiplies by some factor with every new note. In this example, each note tended to have two interpretations so that the number of interpretations doubles with every new note. When we listen to music that has an identifiable beat that we can tap our feet to, such as "Eine kleine Nachtmusik," we rarely have any consciousness of a profusion of possible interpretations. Yet, one can not make any conclusions as to the nature or amount of processing that is taking place in our minds *unconsciously*.

On the other hand, we can make certain conclusions about beat perception based on the nature of computation itself, regardless of the specific algorithm or machine. There are certain universal properties which apply to the computation of *any* problem. One such property is the *order of growth* (Lewis & Papadimitiou 1981) which relates the consumption of some computational resource (usually time or space) with the some measure of the size of the

input. The order of growth of time with respect to the number of notes heard limits the number of rhythmic interpretations that can be pursued.

If one wanted to pursue every rhythmic interpretation as in the example above, the number of interpretations after n beats can be approximated by the equation k^n where k is the average number of ways a single note can be interpreted. For each interpretation that exists, however, a non-zero amount of time must be spent extending it for each new note. Thus the amount of processing time required for the n th note is proportional to k^n . (The amount of space, i.e. memory, required is also proportional to k^n).

Nonetheless, the human beat perception process occurs in real-time, which implies that the amount of processing required to perceive the beat is bounded, and essentially independent of the number of notes heard. Human performance of beat perception does not seem to degrade after hearing “too many” notes! The obvious conclusion is that humans do not pursue all rhythmic interpretations, although one can not argue from this reasoning that we do not pursue hundreds, thousands, or even millions of interpretations, so long as there is some maximum.

This idea may offend the reader’s musical intuition, since we don’t seem to feel that we are pursuing hundreds or thousands of alternate rhythmic interpretations – it seems as if the note values and rhythms are just “there.” We must be skeptical of our ability to gain useful insight into our minds through introspection, however, as it is easy to think of many processes, such as riding a bicycle or recognizing a face, which must involve enormous

amounts of computation to which we have no conscious awareness. As Minsky points out:

In general, we're least aware of what our minds do best.

It's mainly when our other systems start to fail that we engage the special agencies involved with what we call "consciousness." Accordingly, we're more aware of simple processes that don't work well than of complex ones that work flawlessly. This means that we cannot trust our offhand judgments about which of the things we do are simple, and which require complicated machinery.¹⁷

Rhythm Perception vs. Music Perception

Thus far, we have mostly restricted our discussion of music to the time domain, in particular, to the onset times of notes and the inter-onset interval. We have essentially ignored all the other aspects of real performances of music such as pitch, timbre, phrasing, rubato, etc. This restriction is common to rhythm perception research (Povel 1981; Handel & Oshinsky 1981; Longuet-Higgins & Lee 1984; Schloss 1985), and has obvious advantages: one may ignore many of the vastly complicated issues of hearing, pitch perception, voice separation, tonal structure, etc. For precisely these reasons, the model of beat perception proposed in this thesis below considers only note onset and offset timing, and note loudness (see below). Nonetheless, it is worthwhile to address the problematic issues associated with this approach.

I believe that one of the great unanswered questions in the perception of musical rhythm is: In what way and to what extent is the perception of

¹⁷ Minsky (1986), p. 29.

rhythmic structure interdependent on the perception of other aspects of music beyond timing? We have already mentioned the research of Todd (1985) into structural information of systematic timing deviation. Research into the contributions of pitch and rhythm to the perceptual similarity and well-formedness of music phrases has been conducted by Palmer and Krumhansl (1987a & 1987b) and by Monahan and Carterette (1985), but little is currently understood about the direct influences of melody, tonal structure, timbre, or phrasing on the perception of meter.

In order to examine the contributions of some of these aspects of music, in particular, pitch, note duration, and note loudness, I shall begin with an extremely impoverished example in which the only information present is the note onset times, and incrementally augment the piece with different kinds of information. The music is a real performance of a well-known piece keyboard piece performed on a piano by an non-professional musician.¹⁸ Figure 12 below shows the onset times:

¹⁸ Example played by Michael Hawley.

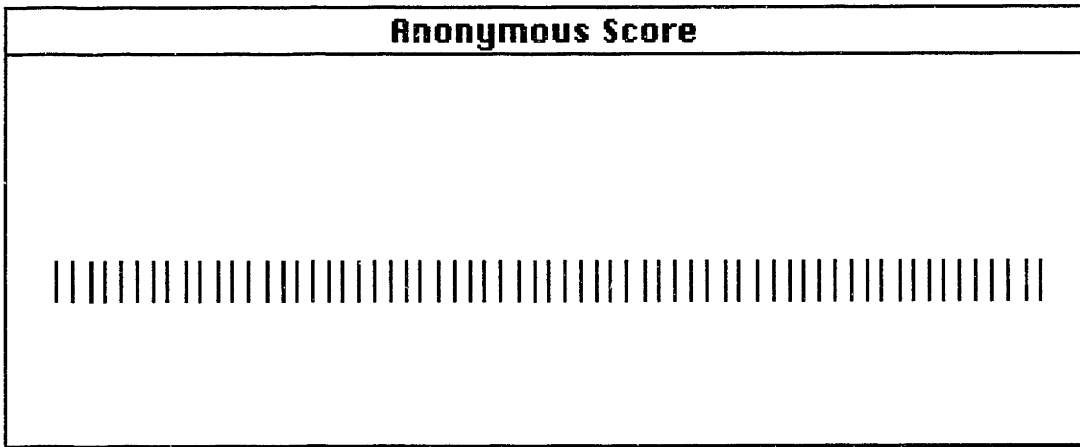


Figure 12.

Although there is an obvious periodicity, no hierarchical organization to the onsets is readily perceivable, either to the eye or ear. When this data is heard in this form, on a single pitch and with a fixed note duration, it simply sounds like the same note repeated with a more or less constant interval. There is no sense of downbeat or measure boundary.

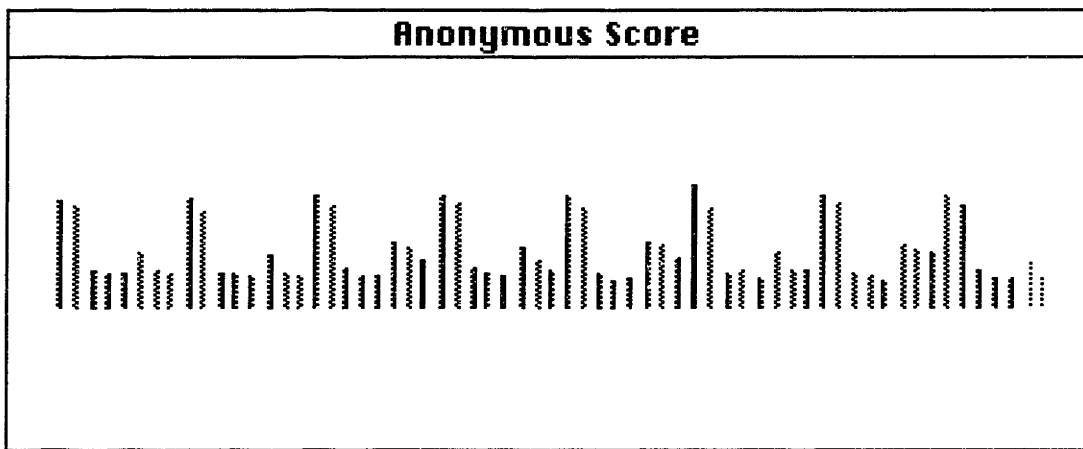


Figure 13.

Figure 13 above shows the same piece, but with the additional information of the note durations (shown as the height of the lines) and the velocity with which the keys were played (shown as the solidness of the lines). Note duration is the length of time a note is held, regardless of when the next note is played, and key velocity is a measure of note loudness.¹⁹

With this information, a pattern of repetition emerges, and one may hazard a guess as to the measure boundaries. There is an observable cycle of eight notes beginning with the first note. The first note of the cycle is nearly always the longest, and it is usually one of the loudest. An educated guess might assume the meter is 4/4, that each measure contains eight notes, and that the bar lines come just before the longest note of the eight-note cycle.

Figure 14 below shows the piece with the pitches included which becomes recognizable as the Prelude I from Book I of *The Well-Tempered Clavier* by J.S. Bach. With the pitch information available, one sees and hears that the piece progresses in pairs of repetitions of eight-note sequences. This additional grouping of pairs induces another level of metrical hierarchy, and we find that our assumption above was slightly wrong. In fact, the notated metrical interpretation is 4/4 where each measure contains sixteen notes, and the measure boundaries come just before the first note of a new pair of repeated eight notes (although many listeners may still choose the original interpretation above). Figure 15 below shows the notation of the first few bars of the piece with the metric hierarchy identified.

¹⁹ This is a simplification, since the notes on a piano also change timbre depending on the key velocity.

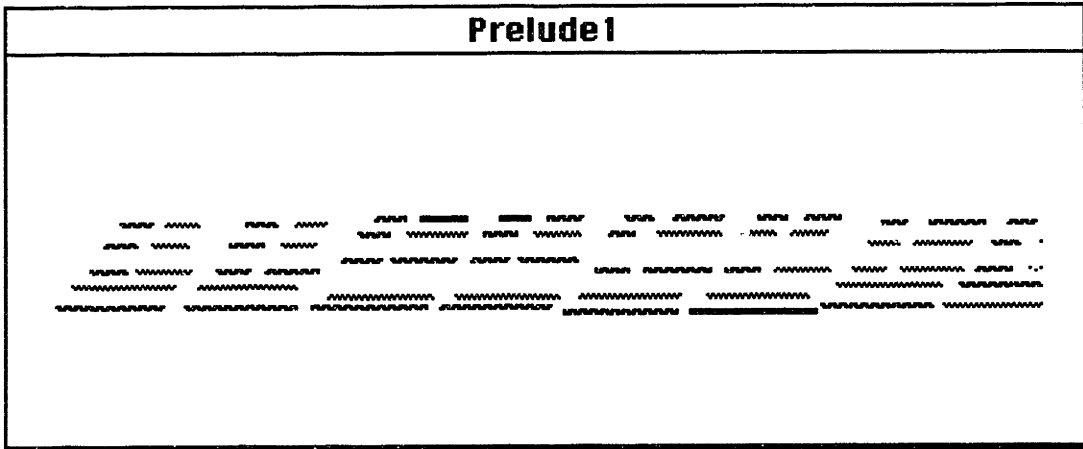


Figure 14.

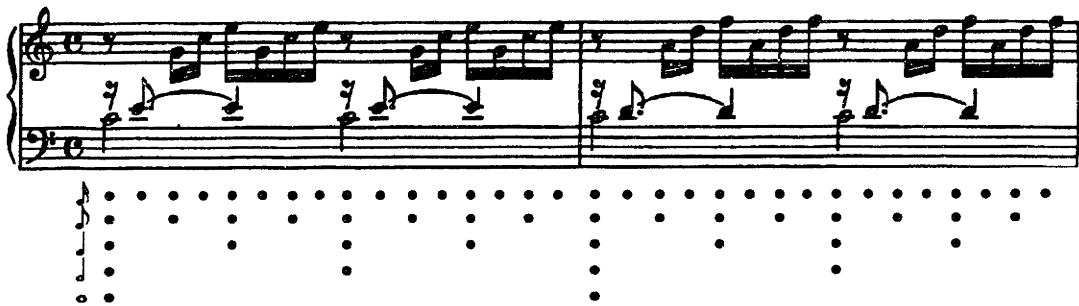


Figure 15.

What is interesting to note about this example is that although most of the metrical structure is perceivable without the pitches, it is impossible to generalize any absolute principles for doing so. The downbeat is recognizable as the longest note in the measure which might lead to the conclusion that relatively long notes should be relatively strong beats (referring to Lerdahl and Jackendoff's definition of "strong" and "weak" beats discussed above). In the same manner, one might expect relatively loud notes to fall on strong

beats, as there are many instances where the stronger beats were played loudly.

While generalizations about strong beats occurring on long and loud notes mostly apply, they are not strictly obeyed. The second eighth-note is held longer than any of the following notes in the measure, yet it falls on a metrically weak beat. In many cases, such as in the second measure, strong beats are played more quietly and for a shorter duration than weak beats.

Furthermore, it is possible to correctly interpret the metrical structure of the piece even when all of the non-pitch cues are absent or contradictory. Figure 16 below shows an unlikely performance of the Prelude in which the durations and velocities were chosen at random (within certain parameters) by a computer.²⁰ Although the result is hardly musical, and probably no pianist would perform the piece as such, the metric structure is still evident. Obviously, a system which ignores pitches, such as the one developed in this thesis, would fail miserably on this input. A human listener, however, can still perceive the beats.

²⁰ The velocities were chosen at random from the approximate range of the original performance. The durations were chosen to be approximately between a sixteenth note and a quarter note.

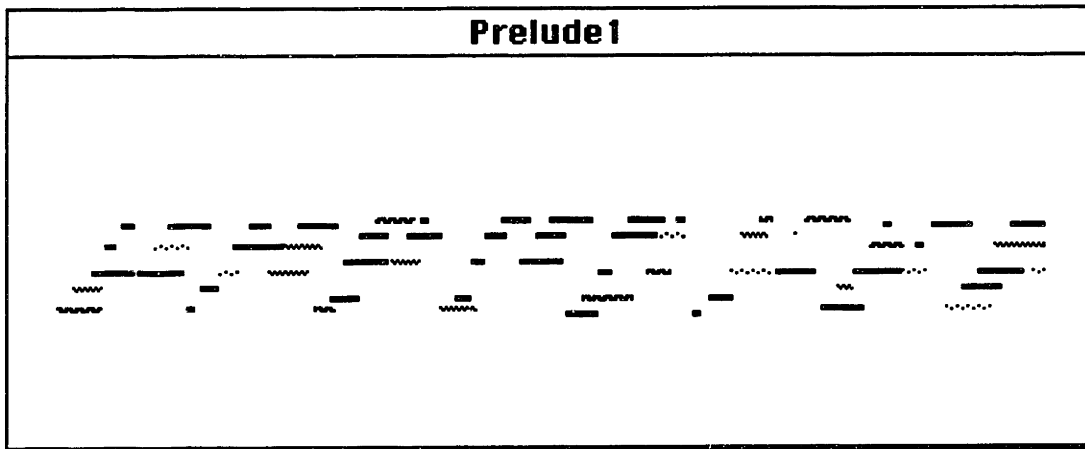


Figure 16.

It may be argued that the correct metrical structure is still perceivable in this case only if listeners are familiar with this particular piece. But in light of the fact that the Prelude maintains its regularity so strictly, I would be surprised if a new listener failed to correctly identify the metrical structure of the random case above after more than a few bars.

The conclusion from this example is that we perceive metrical structure from a variety of sources, which can agree or disagree to varying extents: No single source is always sufficient. It should also be noted that although metrical structure may, in some cases, be perceivable even when certain information is absent, one cannot necessarily conclude that the remaining process is unchanged. Eliminating a rich source of perceptual information such as pitch may force the remaining process to behave in a significantly different manner. One cannot assume that a correct model of the impoverished case is necessarily applicable when all information is considered, especially if the goal is to make conclusions about the nature of human perception.

A complete theory of rhythm perception must include a thorough explanation of the interaction between virtually every aspect of music perception, or to quote the opening sentences of Cooper and Meyer: "To study rhythm is to study all of music. Rhythm both organizes, and is itself organized by, all the elements which create and shape musical processes."²¹

I believe that this lack of simple, orthogonal principles is not particular to music theory, but is common to all attempts at formalism of human mental processes. As Minsky (1986) observes, in the prologue to his book *Society of Mind* :

My explanations rarely go in neat, straight lines from start to end... Instead they're tied in tangled webs. Perhaps the fault is actually mine, for failing to find a tidy base of neatly ordered principles. But I'm inclined to lay the blame upon the nature of the mind: much of its power seems to stem from just the messy ways its agents cross-connect.²²

Metrical Preference Rules

While it would certainly be convenient if one could specify a set of simple rules for obtaining the correct metric interpretation from a performance, the preceding discussion concludes this is not a likely approach. The most detailed attempt at a collection of such rules can be found in the *metrical preference rules* of Lerdahl and Jackendoff (1983). These rules, however, only specify the way in which some given musical parameter contributes to a

²¹ Cooper & Meyer (1960), p.1.

²² Minsky (1986), p. 17.

metrical interpretation. The rules state which of two interpretations would be selected if they differed in only a single dimension.

The first of Lerdahl and Jackendoff's metrical preference rules which are relevant to monophonic music are reproduced below:

MPR 1 (Parallelism) Where two or more groups or parts of groups can be construed as parallel, they preferably receive parallel metrical structure.

MPR 2 (Strong Beat Early) Weakly prefer a metrical structure in which the strongest beat in a group appears relatively early in the group.

MPR 3 (Event) Prefer a metrical structure in which beats of level L_i that coincide with the inception of pitch-events are strong beats of L_i .

MPR 4 (Stress) Prefer a metrical structure in which beats of level L_i that are stressed are strong beats of L_i .

MPR 5 (Length), final version Prefer a metrical structure i which relatively strong beats occur at the inception of notes of either

- a. a relatively long pitch-event,
- b. a relatively long duration of a dynamic,
- c. a relatively long slur,
- d. a relatively long pattern of articulation,
- e. a relatively long duration of a pitch in the relevant levels of the time-span reduction, or
- f. a relatively long duration of a harmony in the relevant levels of the time-span reduction (harmonic rhythm).²³

Each of these rules seems to be perfectly justified, at least in certain musical situations which undoubtedly prompted their formulation. As the findings of

²³ Lerdahl and Jackendoff (1985), p. 75-80.

other researchers do not seem to contradict these rules, one might think to add them as well:

Longuet-Higgins & Lee (1984): Prefer a metrical structure in which there is a minimum of syncopation.

Steedman (1977): Prefer a metrical structure in which the repetitions of melodic fragment occur at a high beat level.

The difficulty with applying this collection of rules to the beat perception problem (aside from the fact that they ignore the timing deviation aspects discussed previously) is that there are no rules about how these rules interact, especially when they conflict and contradict each other. Lerdahl and Jackendoff omit any discussion along these lines. Having the knowledge necessary to choose one interpretation by a single rule above does not generalize into an algorithm for finding the right interpretation when many rules apply. At best, the rules above can serve as a checklist of considerations a good beat perception system should address.

The failure to explain the dynamics of preference interaction is actually indicative of a larger problem associated with the nature of syntactic theories in general. These theories explain how certain structures and features on the musical surface are related to meter, but say nothing about the underlying process of meter perception. Minsky addresses this issue in his paper "Music, Mind and Meaning:"

...this surface taxonomy, however elegant and comprehensive in itself, must yield in the end to a deeper, causal explanation. To understand how memory and process merge in "listening," we will have to learn to use

much more “procedural” descriptions, such as programs that describe how processes proceed.²⁴

The Beat Perception Process

It is important to remember that listeners do not listen to an entire piece of music with pen and score in hand, and then decide upon a metric interpretation afterwards. Rather, beat perception is a real-time, dynamic process which is continuously performed. This distinction is important in light of previous observations: that one’s metric interpretation of a particular sequence of notes generally depends upon the pre-established metrical context. Lee (1985) claims that “the metrical evidence early on in a sequence counts for more than later evidence.”²⁵ The progression of the beat perception process depends upon expectations created by the performance up to a particular moment.

It is not surprising that the metrical preference rules above make no mention of expectation, as expectation is a property of the underlying beat perception process and manifests itself only indirectly on the musical surface. Lee (1985) points out in a paper which partially refutes some of his earlier work:

A general problem with both the proposals in Longuet-Higgins and Lee (1984) and Lerdahl and Jackendoff (1983) is that little or no account is taken of the way in which the listener’s hypotheses about earlier parts of a sequence can affect his hypotheses about later parts.²⁶

²⁴ Minsky (1981), p. 35.

²⁵ Lee (1985), p. 60.

²⁶ Ibid.

The only reference to expectation-like concepts in the metrical preference rules is rule one (parallelism) which, as pointed out earlier, violates Lerdahl and Jackendoff's original premise: that beats and meter are derived entirely from information at the musical surface.

Expectation

Another way of looking at the effect of metrical context on the interpretation of future events is that a product of the beat perception process is an *expectation* that future events will continue to match the current metric interpretation. Depending on the certainty of an interpretation, various amounts of contradictory cues, such as syncopation, are allowable. Recall the example shown above in the discussion of perfectly mechanical performances where a sequence of notes that would probably receive a 3/8 interpretation in isolation could be assigned a 2/4 interpretation if the 2/4 context were firmly established by the preceding bars.

Thus when considering the beat perception process, we must examine both the dynamics of the interaction between different metrical cues as discussed above, as well as the dynamics of the interaction between these cues and existing expectations. These interactions are a product of the *process* of beat perception. A model that does not examine this process cannot possibly account for them.

There is also evidence that expectation plays an even more fundamental role, affecting how rhythmic values are perceived in the first place. In his paper "Categorical Rhythm Perception: An Ecological Perspective," Eric Clarke (1987) demonstrates experimental evidence for the dependency of rhythmic

interpretation on metrical expectation. Clarke conducted an experiment which tested the identification and discrimination of a sequence of two notes whose durational ratios were varied in nine discrete steps between 2:1 and 1:1 while maintaining the same total duration of one beat. The two-note sequences were heard near the end of a short musical phrase which established either a 2/4 (duple) or 6/8 (triple) metrical context: ²⁷



Figure 17.

The ratios of second to last notes in the sequences above labeled "x" and "y" (shown without stems) were varied between 2:1 and 1:1 while the total duration of both notes was maintained at a constant three eighth notes for the 6/8 sequence and one quarter note for the 2/4 sequence.

Clarke found strong evidence that listeners perceive rhythmic ratios categorically; that is, that listeners interpret the ratios of notes as belonging to one of a small number of pre-established categories. In this case, his hypothesis was that 2:1 and 1:1 ratios formed two adjacent categories so that

²⁷ Clarke (1987), p. 23.

perception of ratios between 2:1 and 1:1 would be sharply divided at a specific "break-point."

In the identification experiment, subjects were required to identify each of the sequences as forming either a 2:1 or 1:1 ratio between its two notes. The results pooled from data from both metrical contexts showed that subjects tended to make a sharp transition in judgment at the approximate ratio of 1.3:1. Ratios slightly higher than 1.3:1 were judged as belonging to the 2:1 category by over eighty percent of the responses. Similarly, ratios slightly lower than 1.3:1 were judged as belonging to the 1:1 category by over eighty percent of the responses. The 1.3:1 ratio emerged as a dividing point between two perceptual categories.

The discrimination experiment also strongly supported categorical rhythmic perception. In this experiment, subjects listened to pairs of sequences whose ratios were adjacent steps of the nine variations between 2:1 and 1:1. The pairs of sequences were again presented in context, and subjects were required to identify the order of presentation, i.e. to discriminate which of the pair had the longer or shorter ratio. The results of this experiment showed a strong peak of correct responses for the pair of ratios which straddled the 1.3:1 boundary. This demonstrated that listeners had difficulty discriminating between ratios in the same category, but no trouble when one ratio belonged to the 2:1 category and the other belonged to the 1:1.

Of interest to our discussion on rhythmic expectation, are the results of the experiment when the two metrical contexts, 2/4 and 6/8 are considered separately. Clarke found that while the categorical nature of rhythmic

perception was equally well preserved in both contexts, the boundary between the categories was significantly different. The boundary for the 2/4 category was at approximately 1.35:1, while the boundary for the 6/8 category was approximately 1.2:1. The pre-establishment of either a duple or triple context changed the way listeners categorized the input rhythms. In essence, the metrical context affected the way the rhythms were “heard.”

Disambiguation

The evidence that metrical context affects categorical perception of rhythmic values is particularly relevant to the previous discussion of timing deviation and rhythmic ambiguity. The metric context can be used to contribute to the disambiguation of note values. We can use a theory of the meter as a tool for reducing the number of possibilities by strongly weighting rhythmic interpretations which fit our metrical expectations.

In the “Eine kleine Nachtmusik” example discussed previously, we can use metrical expectation to help choose between interpretations. The note onset times after five notes are reproduced below:



Figure 18.

Assume that the metrical context has somehow been established to be 4/4 and that the first note in the figure above falls where the downbeat is expected. The metrical expectation is that beats will obey a strong-weak alternating pattern, with beats one and three stronger than beats two and four. Below are

reproduced the ten possible rhythmic interpretations from the original example:



Even without considering the “on-timeness” of the note onsets, the correct


interpretation:  emerges as the interpretation which best conforms to the metrical expectations. Beats one and three both have relatively long dotted quarter notes associated with them, and are thus “stronger” than beats two and four which are silent. Furthermore, the eighth-notes fall on beats of the eighth note level, instead of on the sixteenth note level which would not be as preferred. The metrical hierarchy is shown below:



Figure 19.

Expectation can also be used to help disambiguate to the inherent metrical ambiguity presented in the discussion of perfectly mechanical performances

above. In this case, expectation is used in a simpler manner: once a metrical context has been established, it is expected that it will continue.

As a general principle, we can view expectation as a vehicle for disambiguation, the premise being that we hear note values, beats, and meter in a certain way because we expect to hear them that way, and that if our expectations are not too badly contradicted, they will be continued. Narrowing the space of rhythmic and metric interpretations by mainly considering interpretations which meet existing expectations provides a strategy for managing the plethora of possibilities.

Beat Perception of Perfectly Mechanical and Real Performances Revisited

I would like to return to the previous discussion of perfectly mechanical and real performances, and particularly to my earlier claim that recovering intended note values (i.e. the perfectly mechanical version) of a real performance is not an independent sub-problem of beat perception. Many approaches to beat perception (Longuet-Higgins & Lee 1984; Lee 1985; Lerdahl & Jackendoff 1983; Steedman 1977) consider only perfectly mechanical input, which limits the direct application of their theories. Theories of beat perception of perfectly mechanical input may provide many useful insights, but it is wrong to assume that beat perception proceeds by first converting real performances to perfectly mechanical ones.

Clarke's findings indicate that rhythmic interpretation actually *depends* on the metrical context. If this is true, then the conversion of real to perfectly mechanical performances cannot take place independently. Furthermore, it can be shown that it is not generally necessary to perceive the intended

rhythmic interpretation in order to perceive the intended metric interpretation.

For example, consider the case where a performer, intending to play a triplet, actually performs the three notes very imprecisely such that half of the audience perceives a triplet and the other half perceives an eighth and two sixteenth notes. Although some listeners perceive a different rhythmic interpretation than the others, this does not, in this case, affect their metric interpretations. Everyone will continue to tap their feet the same way. Metric and rhythmic interpretations must be compatible, in the sense that the values of the notes comprising a beat at a particular beat level must add up to the length of the beat, but the exact rhythmic interpretation is not important. In our example, it is only necessary that listeners perceive three notes which fit into the time of one beat; it does not particularly matter what their specific values are.

This reasoning implies that rhythmic interpretation and metric interpretation proceed simultaneously as inseparable parts of the same process. If this is the case, then understanding beat perception of perfectly mechanical performances is not very useful as it only begs the question of obtaining the perfectly mechanical performance in the first place.

When considered as a dynamic process we see that beat perception involves many elements such as expectation and the interaction between rhythmic and metric interpretation which do not manifest themselves directly in the syntactical surface of the music. A model which considers only syntactic structures and preference rules will fail to account for these interactions.

Summary

- In general, a given performance of music has a large number of ambiguous metric interpretations.
- There are two sources of ambiguity: rhythmic ambiguity create by timing deviations, and the inherent metrical ambiguity present even in the absence of rhythmic ambiguity.
- We prefer interpretations that are unsyncopated, whose long, loud notes fall on important beats, and whose timing deviations are systematic and correspond with the underlying musical structure.
- Expectations created by metrical context is a powerful tool for disambiguation.
- Rhythmic interpretations depend on the metrical context.
- We must examine beat perception as a dynamic process in order to understand the interaction between the various elements that influence our perception. The problem cannot be solved by examining the syntactic musical surface.

Previous Work in Beat Perception Machines

There are surprisingly few examples of beat perception algorithms that have been implemented on a computer. Although beats and meter have been the subject of a large body of theoretical discourse, a large gap exists between the theories and their implementation. On the other hand, the programs written have tended to focus on a small set of idealized input which, as discussed earlier, may not yield findings transferable to more general cases. Nonetheless, computer implementation of theories is an extremely important and useful explorational tool.

An implementation can be regarded as a sort of "acid test" of a theory. The concrete expression of a theory as an algorithm on a computer forces the theorist to consider many problems and assumptions which a theory on paper may not address. Additionally, skeptics of the theory can test the performance of the program in the laboratory under many conditions.

Kinds of Theories

The nature of the previous work in beat perception machines is strongly influenced by the original goal behind the work. While certain systems are implemented as a test of a theory, or as a medium for understanding a theory's details, others are constructed for more practical reasons such as automatic transcription, and may not possess a formal theoretical basis.

The implementations of theoretical works can also differ with regard to the intent of the theory. Some theories such as that Lerdahl and Jackendoff are mainly concerned with properties of musical structure, while others take a

more cognitive approach, with the intention of modelling the process of human perception. The difference between these approaches is analogous to the syntactic/semantic distinction. Minsky (1981) draws an analogy to linguistics where one can regard a grammar book as a theory of the properties of sentences which does not address questions of underlying human perception, such as which sentences are meaningful.

This thesis is primarily interested in work which presents *perceptual theories* of beat perception: theories which attempt to explain the way the humans perform the process of beat perception, and why this process may function the way it does.

Constraints on Perceptual Theories

Perceptual theories always possess certain constraints to which non-perceptual theories are not bound. Because human beat perception occurs in real-time, computer implementations of perceptual theories must be expressed as a real-time algorithm. As discussed previously, real-time in this sense says nothing about the amount of computation that can be expended on the problem, but does constrain the way the amount of computation can *grow* with respect to the size of the input. A real-time algorithm may take days or weeks to complete processing on even a few notes, but the maximum processing time for a given note must be bounded.

Another property of an implementation of a perceptual theory is that the algorithm must receive its input sequentially, as a human listener hears music. The algorithm cannot produce a metric interpretation by analyzing large chunks of the score and deciding upon the beats and meter at a later

time. Instead, it must proceed incrementally, yielding its results (i.e. tapping its foot) “on the fly” as a human does.

There is a further constraint on perceptual algorithms imposed by the nature of human memory²⁸. Research on human musical memory indicates that there are severe limitations on how much “raw” auditory data is retained over time. Deutsch (1982) has found that recall of absolute pitch information is affected not as much by gaps of time, however, but by the presence of new pitch material, especially when the new material is close in pitch (within a whole tone) to the pitches tested. In the rhythm domain, Essens and Povel (1985) found that reproduction of repeated rhythmic patterns less than three seconds long was poor unless the patterns could be represented as a hierarchical (or metrical) structure.

The essential finding in this literature is that the precise human recall of auditory information, if such a thing actually exists, is extremely limited. The relevance to beat perception machines is that a perceptual model should not depend on precise recall of past musical input beyond some small amount of time. The evidence indicates that recall of musical information is highly dependent on how well the auditory signal can be “parsed” into musical structures. The “on the fly” nature of a perceptual algorithm’s output must also reflect an “on the fly” treatment of its input.

²⁸ A thorough review on the research in human memory of music can be found in Sloboda (1985).

Type of Input

Another way of characterizing a beat perception algorithm is by the type of input the algorithm is designed to work with. The work in automatic transcription by Chowning et. al. (1984) and Schloss (1985) accepts an acoustical signal as input. However, their metric interpretation mechanisms operate on higher level structures generated from signal processing analysis. As for the implementations of perceptual theories, however, all of the systems that the author is aware of accept only perfectly mechanical input. In some cases the input has been augmented with note loudness information, and in others, pitch information has been removed. But in general, timing deviations are always eliminated.

As stated previously, the main problem with beat perception models of perfectly mechanical input is that it is a fallacy to assume that humans convert real performances to perfectly mechanical ones as a first step. If the processes of obtaining the rhythmic interpretation and metric interpretation are indeed parts of the same process, as argued above, then solving the seemingly easier problem of accepting only perfectly mechanical input may not provide much progress towards solving the problem in general.

Review of Previous Work

The various implementations of beat perception systems on a computer that are discussed in this thesis are show below, according to the goal of the system:

Perception Modeling:
Steedman, 1977
Longuet-Higgins & Lee, 1984

Lee, 1985

Live beat tracking of jazz performances:
Dannenberg & Mont-Reynaud, 1987

Automatic Transcription:
Chowning et. al. 1984
Schloss, 1985

Longuet-Higgins and Lee (1984)

Longuet-Higgins and Lee's (1984) work has been discussed in some detail in various sections above. While the work provides an interesting basis for perception of meter from note durations alone, the approach is fairly simplistic, and can account for only a narrow range of unsyncopated music. In addition to the problems inherent with any system which only accepts perfectly mechanical input, this approach also fails to model any sort of rhythmic expectation.²⁹

Steedman (1977)

Steedman's (1977) approach narrows in on one factor in metrical grouping, namely, the contribution of melodic repetition. His algorithm discovers higher level organization assuming that the beats and perhaps some metrical grouping have already been identified. This is an interesting approach which is complementary to the model proposed in this thesis. Steedman's model would discover the melodic repetition in the *Well-Tempered Clavier* example from above, whereas a model which examines only the time domain would not. Steedman's work is especially notable in that it appears to be the only model of melodic contribution to meter.

²⁹ Further criticisms of this approach can be found in Lee (1985).

Dannenberg and Mont-Reynaud (1978)

Dannenberg and Mont-Reynaud's (1978) program is interesting as it appears to be the only attempt so far at performing live beat tracking. Their system was designed as part of a system for following live jazz improvisations, and thus it accepts real performances of music in the form of MIDI. Their approach is to attempt to track repetitions of a single time interval. The algorithm functions in two modes. At first, it examines the input until it finds a succession of three "healthy" notes whose inter-onset intervals are roughly equivalent. Once this condition is satisfied, the program attempts to track the same inter-onset interval by treating new "healthy" notes which occur more or less at the appropriate interval as evidence towards a changing tempo.

This approach is fairly simplistic, presumably due, in part, to the computational constraints imposed by a live situation. The system has no concept of downbeat or meter, nor any form of higher level expectation. For example, if it were tracking eighth-notes, it would not handle input with triplets correctly.

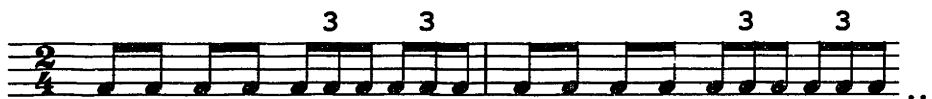


Figure 20.

If their system were presented with the above sequence, it would consider every note to be an eighth-note and continuously redefine its beat value without ever settling into the correct, constant rhythm.

Indeed, Dannenberg states in personal correspondence that “the results were pretty disappointing all around.”³⁰ The preceding discussion on the importance of higher level structure and expectation can explain many of their difficulties.

Chowning et. al. (1984) and Schloss (1985)

The work in automatic transcription algorithm by Chowning et. al. (1984) and Schloss (1985) does not make any attempt to model human perception, and is thus the farthest from perceptual theory of the work discussed. Their algorithm is not restricted by any of the constraints placed on perceptual algorithms; it can perform any sort of computation over any range of the piece of music in question without any regard to sequentiality or order of growth. The first step in Schloss’ system, for example, is to compute a histogram of inter-onset intervals over all the notes in the piece.

Given these computational advantages, one might think that their approach must be superior, but this is not necessarily the case. Human perception of rhythm is the result of a complex process with many interactions. The way we “hear” a piece of music is affected as much by this underlying process as by the collection of notes, timbres, and times that make up the piece. While many musical features such as common time intervals are evident on the musical

³⁰ Dannenberg (1988).

surface, many of important musical structures are not. Statistical approaches will never recover these structures completely. Indeed, in the skill of transcription, human ability far outstrips any synthetic system, and there are no systems in existence which can accurately transcribe even moderately complex music.

C.S. Lee (1985)

The previous work most relevant to the model presented in this thesis is a perceptual theory proposed by C.S. Lee (1985). Lee extends the work of Longuet-Higgins and Lee (1984) with a model which accounts for higher level expectations induced “bottom-up” from the data. This principle is also used in the model presented in this thesis, although with different criteria for the induction. Like Longuet-Higgins and Lee (1984), Lee’s model is concerned only with the note durations of perfectly mechanical, monophonic music.

The essential idea behind Lee’s approach is that listeners establish metrical units of time by finding two adjacent, equal units of time initiated by notes that are no shorter than any of the other notes within the units. Once a metrical unit is established, higher level units are created by the same principle whenever a note on a beat is encountered that outlasts the current unit in duration. His example is reproduced below:³¹

³¹ Lee (1985), p. 64.



Figure 21.

The listener begins by finding the metrical unit of one quarter-note beginning with the first note. This satisfies the requirement that the first notes (the dotted eighth and dotted quarter) of the first two adjacent metrical units are not shorter than the notes within. A metrical unit of a dotted eighth-note beginning with the first note would not satisfy this requirement, however, as the sixteenth-note would be shorter than the dotted quarter-note. The resulting breakup of the notes into quarter-note metrical units is shown below:



Figure 22.

The next step is to try to find higher level metrical units which are induced when a note on the quarter note beat above lasts longer than a quarter-note. The first such note is the dotted eighth-note (shown above as a quarter tied to an eighth) on the second beat.

The process of finding a higher metrical unit is accomplished in the same manner as above, except that we only consider notes that fall on the quarter note beat established above. We look for a unit that is an integral number of beats, starting with the dotted quarter, and then choose the smallest unit that

satisfies the condition that the first notes of the first two units are not shorter than any other notes on the beat. A metrical unit of two beats does not satisfy these conditions because the first note of the second two-beat unit would be an eighth-note, which is shorter than the dotted quarter note on the next beat. The metrical unit of three beats, however, does satisfy the constraints, and thus becomes the higher level interpretation shown below:



Figure 23.

Lee's model is too simplistic to regard as complete. He states himself that "the proposal ... is a speculative (and somewhat) partial attempt to account for the way in which the listener interprets a particular sequence as the realisation of a particular rhythm"³², and it is trivial to formulate an example on which his model fails:



Figure 24.

In this example, Lee's algorithm would find the incorrect metric unit of a dotted eighth-note, leading to a 3/8 interpretation:

³² Lee (1985), p. 67.



Figure 25.

Once the dotted eighth-note unit is found, there is no opportunity for recovery. A more natural interpretation of this sequence would be in 2/4:

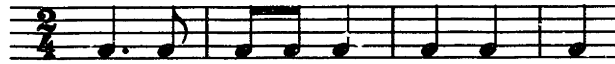


Figure 26.

Nonetheless, Lee addresses certain issues that other theories discussed here ignore, namely, the induction of higher level metrical units, and the importance of metrical expectation. These concepts are central to the model presented in this thesis.

An Agency for the Perception of Musical Beats

Society of (Musical) Mind

The work in this thesis is strongly influenced by Marvin Minsky's ideas derived primarily from two sources: the book, *The Society of Mind* (Minsky 1986) and the article, "Music, Mind, and Meaning" (Minsky 1981). The following is a brief discussion of the general principles of Minsky's ideas as they pertain to music and beat perception.

Agents and Agencies

Minsky's main tenet is that the human mind derives intelligence from a vast number of small, unintelligent processes that are organized in very special ways. These processes when viewed as single units are known as *agents*, whereas collections of processes which work together towards a common functionality are known as *agencies*. Although a single agent may seem to do nothing more intelligent than respond in a straight-forward manner to a limited number of input conditions, thousands of agents properly configured into an agency, can perform the immensely complex tasks of playing a violin or writing a symphony.

Agencies are constructed by organizing agents into hierarchies where the flow of information is essentially limited to communication between an agent's superior and subordinates. Furthermore, because individual agents are such simple machines and can not possibly possess any sort of sophisticated language, communication is restricted to simple "switch-throwing" and sensing protocols, such as activating subordinate agencies or informing a superior agent of success.

Our minds thus contain hundreds, perhaps thousands, of agencies for listening to, performing, and composing music. Each of these agencies use many common sub-agencies such as the agencies which hear pitch and timbre, the agencies that follow melody and harmony, and the agencies that hear beats and rhythms. This thesis focuses on a model of the agency that perceives beats, and how a beat perception agency might fit into the larger agencies that process and perceive other musical structure.

A useful way of thinking about the organization of agents is the concept of B-brains. The idea is that we have an A-brain which is a thinking, intelligent set of agencies which are directly connected to our sensory and motor mechanisms, but we also have a B-brain which is not. The B-brain is also a set of thinking, intelligent agencies, but one whose inputs and outputs are connected only to the A-brain. This arrangement allows the B-brain to have a kind of global perspective on the A-brain's activities that the A-brain would be incapable of.

This principle provides a good computational model for the kinds of musical processing involved in noticing repetitions and variations of phrases. While an A-brain processes the sounds, a B-brain can observe the A-brain's sequence of processing and notice when the same or similar processing has occurred. Thus, the B-brain can be aware when sequences of rhythms, pitches, or timbres are repeated. Certainly there is also a C-brain which monitors the B-brain and can detect more abstract similarities such as variations on a motif or the relationships of melodies to harmonies and counter-melodies. How long this chain of "X-brains" goes determines the levels of abstraction that humans

are capable of perceiving. When people compare symphonies to paintings, or poems to sculpture, perhaps they are seeing analogies in a brain that is many levels deep.

What Are Beats and Meter Good For?

As stated previously, the goal behind my thesis work is to formulate theories not about the syntactic structure of music, but about the *process* of human perception of music. To this end, this thesis will address how a collection of agents can be organized into a beat perception agency, and how this agency might be organized in relation to the other agencies that participate in the diverse tasks of music perception. The first step in thinking about beat perception is to assemble some ideas about why we even perceive beats and meter at all. What are they good for?

Minsky suggests that music may be a sort of playground for learning about time. He asks, "How on earth does one learn about time? Can one time fit inside another? Can two of them go side by side? In music, we find out!"³³ In personal conversation, Minsky has expanded upon these ideas, suggesting that as we listen to rhythms, we are constantly trying to match chunks of time delineated by musical events. When we find two adjacent chunks which match, we group the two chunks into a larger chunk. The larger chunk represents the original chunks abstractly as two of the same thing.

One of the most obvious uses for such grouping is that it provides a flexible mechanism for representing musical time. Instead of storing two measures as

³³ Minsky (1981), p. 34.

one continuous structure, the measures can be broken into halves that are expressed as two things which are the same length of time. The individual measures, of course, can also be broken down into beats, the beats into subdivided beats, and so forth. This hierarchy allows the timing of musical events to be associated with the various levels of beats, instead of with an absolute scale such as milliseconds. The time of a note onset can be represented as "the third beat of the measure" instead of as "2,321 milliseconds." The advantage to representing time in terms of metrical hierarchy is that timing becomes relative to the musical source, essentially eliminating the importance of tempo. Thus the same phrase played at significantly different tempos can still be easily recognized. (Obviously, *some* tempo information is retained, as we generally know when a phrase is played too slowly or too quickly).

Additionally, representing musical time in a metrical hierarchy can assist in musical performance. The performance processes can utilize the hierarchy of beat levels to assist the sequencing of the agents involved with the control of motor activities. The performance agents can be organized according to the metrical hierarchy, with a different class of playing agent for each beat level, each responsible for the time span of one beat. In this arrangement, each beat level playing agent only has to be responsible for activating two or three lower level agents at the right times. The lower level agents behave identically until there are no lower level agents, at which time the agents responsible for moving the appropriate muscles are activated and notes are played. This is undoubtedly a far too simplistic model, as the planning involved with our motor facilities is fantastically complex. However, beat levels can provide hierarchical input to the planning agents.

As a generalization of the metric preference rules discussed above, one might say that important musical events tend to be on downbeats, or, in terms of a metrical hierarchy: *important musical events tend to be beats of higher beat levels*. This phenomenon makes sense in the context of the above discussions, both for perception and performance. Beats that are perceptually important serve as anchor points to the beat perception system which enables the other musical events to be properly structured. In performance, important notes are more accessible to the higher level planning agencies, and are more likely to be performed correctly. The correct performance of the important notes can serve as a higher level goal.

A Beat Perception Agency

The discussion above suggests the form for the model of beat perception presented in this thesis. This model separates the beat perception process into two parts: an agency for determining the importance of notes, and an agency which divides the music into hierarchical chunks of time based on the periodicity of important notes. The main thrust of the work in this thesis is on the latter agency, the one involved with finding the metrical hierarchy once note importance has been assigned. A brief discussion of how the note importance agency might function is given, followed by a detailed explanation of the metrical hierarchy agency.

The description of the model is somewhat involved; it cannot be meaningfully formulated in terms of one or two simple principles. This is not surprising considering the complex issues involved with beat perception

discussed throughout this thesis. This section is intended to explain, in a sufficient amount of detail, the essential workings of the model. The following section describes the implementation of this model on a computer, and shows several examples of its operation.

Musical Input

The model presented below accepts as input, real performances of monophonic music represented at the “note level” (as opposed to an acoustic signal), such as that produced by MIDI instruments. While the model should function properly with any genre of metrical music as defined earlier, the model was developed with Western classical music as the focus. The melodies from the opening bars of Mozart’s piano sonatas represent typical examples.

Although the model allows significant timing deviation (30% in this implementation), the current model does not track tempo. It expects that a quarter note at the beginning of a piece will have more or less the same time value at the end. Thus, the musical input must be played with a fairly constant tempo. This limitation is not believed to be a difficult one to overcome, and a scheme for extending the model to track tempo is given in a later discussion.

The Note Importance Agency

The first step of the beat perception process is to assign a measure of importance to each note as it is heard. The agency which performs this task is influenced by virtually every aspect of music that can affect beat perception. One can think of this agency as the union of several independent agencies

each of which examines a single aspect such as note duration, phrasing, repetition of melody, note loudness, harmonic structure, etc. Some of these agencies, such as those that observe repetitions or parallel structures and those that deal with tonal structure, are deeply involved with the music perception process as a whole. Their contributions to the beat perception agency may only comprise a minor part of their function. Many of these agencies actually depend on the beat perception agency in order to function.

An agency which notices repetitions of melodies, for example, must use the beat perception agency to structure the candidate melodies for matching, yet its results are used as input to the beat perception agency. This kind of circularity provides a natural explanation for the persistence of metrical contexts. Once a context is established, the circular dependencies have a tendency to maintain the context, even if local evidence is contradictory. The circularity gives the process momentum by providing opportunities for the note importance agencies to continue to find important notes that meet the metrical expectations.

One of the problems with circular dependencies is that the process needs to start somewhere. If a process were exclusively dependent on itself, it would be impossible to initiate. Fortunately, there are large parts of the note importance agency which do not depend on the beat perception agency. Until the beat perception process has found the beat, the processes which depend on it must be dormant. Thus, at the beginning of a piece, before a metrical context can be inferred, the source of information for assigning note importance must come from the musical surface. Note duration, loudness,

and timing must make up the primary input to the beat perception process before a beat is found.

As mentioned above, the main focus of this model is the agency which finds the hierarchical periodicity of notes once the importance has been assigned. In the implementation of the model described below, note duration is used as a crude substitute for importance.

The Metrical Hierarchy Agency

Assuming that the process for assigning note importance functions properly, an agency for discovering the metrical hierarchy from the periodicity of important notes must be constructed. The hierarchy found by this agency should have the property that, for the most part, the more important notes should be beats at higher beat levels.

The metrical hierarchy agency is constructed from hundreds of simple agents called *beat-levels* which are connected to each other in special ways. Each of these beat-level agents represents a different theory as to which notes constitute beats. Each beat-level attempts to fit new notes into its theory, and maintains a measure of how well it is performing. The principle of this model is that the beat-levels which perform well will correspond to the beats and metrical hierarchy which human listeners perceive.

Beat-level Agents

A beat-level agent identifies notes that are on the beat by matching adjacent chunks of time which make up an integral number of beats. We define a

chunk of time as the time delineated by two (not necessarily consecutive) notes. The notes which delineate matched chunks are on the beat.

When a beat-level is created, it has an *initial chunk*, which is made up of the last heard note and some previous note. The length of the initial chunk shall be referred to as the *beat interval*, and is limited to some maximum and some minimum which correspond to the longest and shortest perceivable beats. These figures are generally placed at 2 to 3 seconds for the maximum, and 150 to 180 milliseconds for the minimum (Fraisse 1982). For the implementation of this model, the more conservative figures of 1.5 seconds and 280 milliseconds were chosen, which correspond to the range of tick produced by a typical metronome (Rowe 1989).

Figure 27 below depicts a newly created beat-level. The vertical lines show the note onset times of the rhythmic sequence notated in figure 28

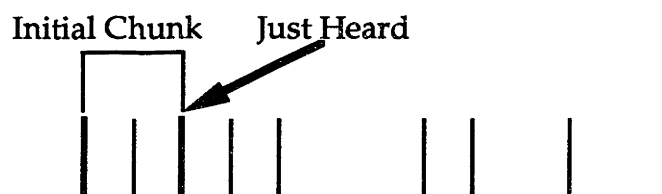


Figure 27.



Figure 28.

The beat-level agent maintains a *current chunk* which is formed by the second note of the original chunk and the note just heard:

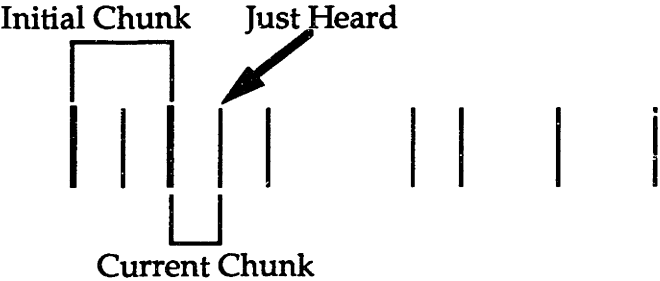


Figure 29.

The beat-level matches the current chunk against the original chunk. The current chunk matches if its time is approximately an integral multiple of the beat interval. The *tolerance* of the match is defined as the percentage that the length of the current chunk can deviate from the ideal interval. (For the implementation of this model, 30% was used). Figure 30 shows a current chunk which matches with an integral multiple of 1 (i.e. the current chunk is approximately the same length as the original):

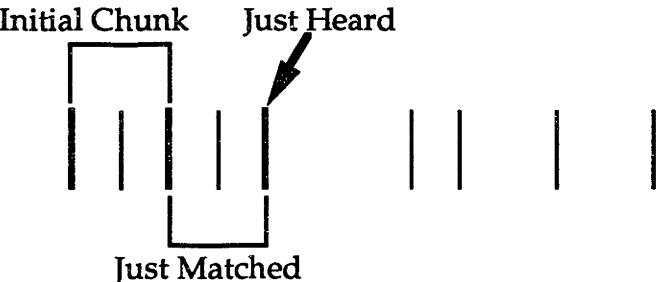


Figure 30.

Once a match is found, the current chunk is moved so that its first note is the second note of the chunk just matched. Then the process continues:

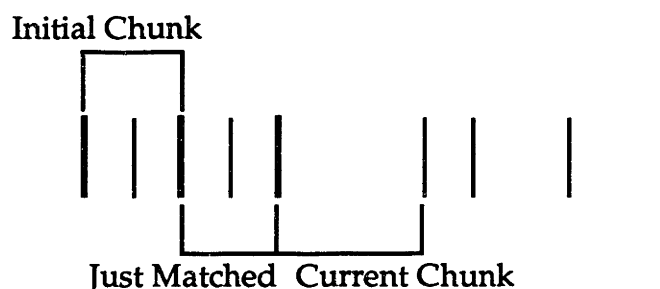


Figure 31.

Profusion of Beat-levels

We know that there are significant timing deviations in real performances so that the beat-level agent must allow for some range of deviation. In this case, there may be more than one note which would form a matching current chunk. On the other hand, even when a match can be found, the correct interpretation may be to ignore the match, i.e. the note which forms the matching chunk may not be on a beat. As an example of the second case, shown in Figure 31 above, the current chunk should not match the original chunk, as the note just heard is on the up-beat (although depending on the match tolerance, the beat-level may consider the current chunk to be a valid match).

Thus, with every note heard, there are potentially a number of ambiguous choices, as discussed earlier in this document. How should the beat-level agent handle this ambiguity?

In a "Society of Mind" model, it is extremely important that no single agent possess "too much" intelligence. For the most part, each agent must interact only with its superior and subordinate agents, and each agent can only have access to a limited amount of information and memory. We have defined the beat-level agent's functionality so that it is only concerned with finding beats. It does not have access to, nor the ability to utilize, the information that could enable it to make a choice between ambiguous possibilities. Those decisions must be made at a higher level.

If the beat-level agent cannot choose between alternative actions, it must proceed by taking *all* of the alternatives. This is not difficult to accomplish. Every time a match can be made, the beat-level replicates itself, creating another beat-level. One of these levels performs the match, and the other does not. One beat-level will then be guaranteed to have made the correct choice, but the decision as to which will be deferred to a higher level agent and a later time.

Thus, as a beat-level proceeds, it generates more beat-levels, each with a different version of which notes are on the beat. The newly generated beat-levels will produce even more beat-levels as new notes are heard, however, as discussed previously, the number of beat-levels cannot be allowed to grow indefinitely.

Scoring Beat-levels

Limiting the number of beat-levels is accomplished by assigning each beat-level a score, and allowing only some number of the highest scored beat-levels to continue. The score is calculated from three parameters: the

importance of the notes on the beat, the size of the timing deviations for the notes on the beat, and the number of syncopations. The better scores should be given to beat-levels that found more important notes, smaller timing deviations, and fewer syncopations.

It is interesting to note that all three parameters that contribute to the score can be calculated for each chunk of time that is matched (i.e. for each beat that is found). The importance of the note on the beat is provided by the note importance agency described above, and the timing deviation is easily calculated from the length of the current chunk and the beat interval. The number of syncopations is defined as the number of times that a beat does not have a note on it. This is the case every time the current chunk matches a multiple of the beat interval greater than one.

Therefore, a beat-level can be scored by calculating a score for each chunk that is matched. The score for a chunk can be thought of as a measure of how well formed the chunk is. A *perfectly well formed* chunk can be defined as a chunk that is precisely one beat interval long, and whose delineating notes were assigned a greater importance than all of its interior notes. (The exact algorithm used for calculating the score of a chunk will be discussed in the next section on the implementation of the model).

Reconverging Beat-Levels

There is one case where a beat-level does make an immediate choice. This occurs when matching a chunk would result in two beat-levels that were essentially identical. Consider the following example, where a beat-level has just found a beat:

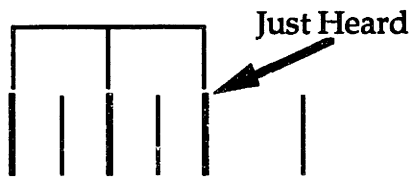


Figure 32.

A beat-level must also exist which is identical in all respects to the above beat-level except that it did not identify the most recently heard note as a beat:

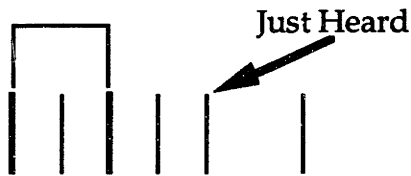


Figure 33.

When the next note is heard, however, there is a potential problem. If both of the above beat-levels identify the note as a beat, the results would be as shown in Figures 34 and 35 below:

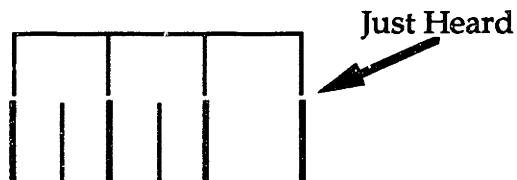


Figure 34.

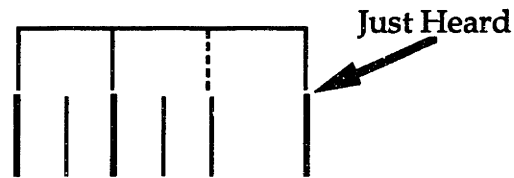


Figure 35.

These beat-levels are essentially equivalent. The objective behind beat-level replication is to explore every alternative; nothing is gained when two replications reconverge on the same interpretation. On the contrary, reconvergence results in a disastrous amount of redundant computation. Beat-levels that have reconverged will behave identically to future notes, consuming twice the computational resources. The problem is that they will receive virtually equal scores so that they will be indistinguishable to the agencies (discussed below) which limit the number of beat-levels. If the identical beat-levels are performing well, the limiting agencies will interpret them as two good interpretations which are worth pursuing, not realizing that one is redundant. If beat-levels are allowed to reconverge, computational resources will be taken away from alternatives. In the extreme case, a beat-level can reconverge so many times that all other beat-levels are eliminated.

Fortunately, preventing reconvergence is not difficult. The mechanism employed is similar to Minsky's (1985) scheme of *cross-exclusion*, where agents are connected so that the activation of one inhibits the others. In our case, all the beat-levels which were generated directly or indirectly from the same original beat-level are connected together. This group shall be referred to as a *family*. When a beat-level identifies a note as a beat, it inhibits all lower scoring members of its family from matching on the same note.

match a chunk at the same time. As an example, in the figure above, the current chunk is exactly 1.5 times the lower level beat interval. If the tolerance for deviation were great enough for the lower level to match the current chunk, there is some ambiguity as to whether it should match this chunk as one or two beat intervals. The presence of the higher level, however, forces the lower level to match the current chunk as two beat-levels because the higher level matches the current chunk as one higher beat interval, and the lower level subdivides the higher level by two. Thus, the two beat-levels will always be "in sync." This is shown in figure 37 below:

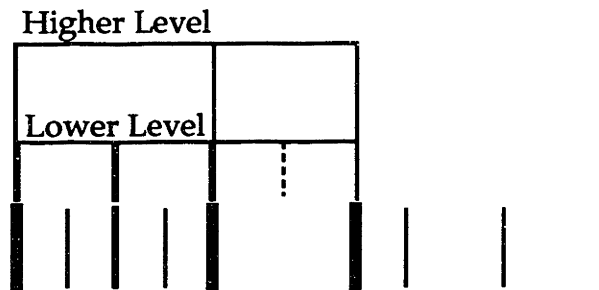


Figure 37.

Superior, Subordinate and Top-Level Meter Agencies

Thus, the higher level of a meter agency serves as the superior agent, and the lower level the subordinate. The lower level, however, can itself be a meter agency, if it is connected to an even lower level whose beat interval is an integral divisor of the lower level's. In general, a meter agency is comprised of a superior higher level and a subordinate meter agency.

For the current example, the lower level in figure 37 above can serve as a higher level to a beat-level whose beat interval is half the lower level's. The resulting meter agency is shown in figure 38 below:

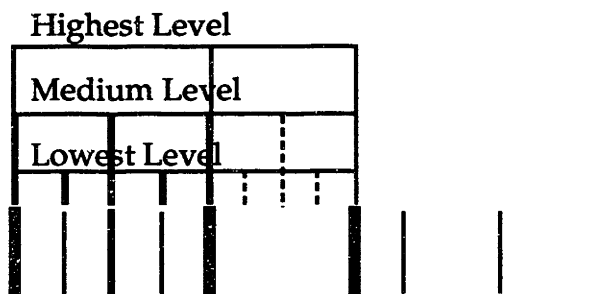


Figure 38

The relation between superior and subordinate is always relative to the particular meter agency in question. The medium level in figure 38 above is the superior in the meter agency that contains itself and the lowest level, however, it is the subordinate meter agency to the highest level in the meter agency that contains all three. This relationship is show pictorially below:

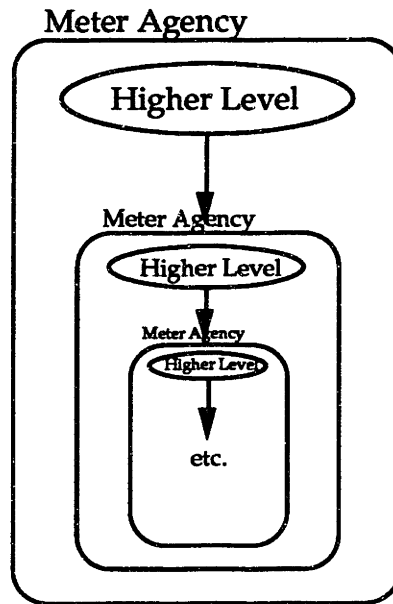


Figure 39.

When a meter agency does not serve as a subordinate to any other agent, it is called a *top-level* meter agency. The outermost box in the figure above represents a top-level meter agency.

As discussed above, whenever a beat-level is faced with a choice of action, it always replicates itself, resulting in one agent that took the action and another that did not. When beat-levels are connected into meter agencies, their behavior is unchanged in this respect, however, the replication occurs for every beat-level. The entire meter agency is copied.

Scoring Meter Agencies

The score of a meter agency is the sum of the score of the higher level and the score of the subordinate meter agency. Thus the score of a meter agency depends on the performance of all of its beat-levels.

The best scores will be given to meter agencies which have, on the whole, the most well formed chunks at all levels. In figure 40 below, the meter agency shown gets a high score because each chunk at each level is well formed (numbers indicate the importance of each beat). Even though there are some missing beats, every chunk's interior notes are less important than its delineating notes. Additionally, each chunk is exactly an integral multiple of the respective beat-level's beat interval.

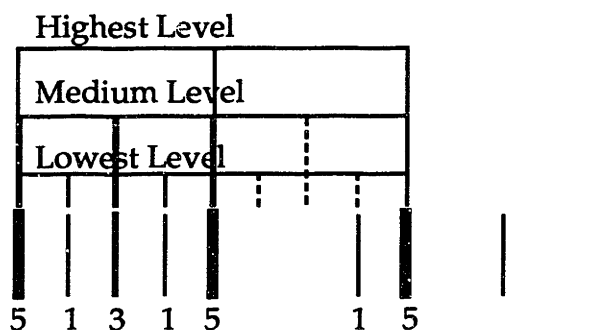


Figure 40.

Creating Meter Agencies

So far, nothing has been said about the process of how meter agencies are created in the first place. This process is of extreme importance because it models the way expectations are created as the music is heard sequentially. Recall the example from the discussion on perfectly mechanical music where a sequence which would most likely be interpreted in 3/8 in isolation, could receive a metrical interpretation in 2/4 if a strong 2/4 context were established by the preceding bars:



Figure 41.



Figure 42.

For the 2/4 interpretation to emerge, meter agencies must be created which expect a duple subdivision of measures into quarter-notes and quarter-notes into eighth-notes. These agencies must be created and established before the third bar. On the other hand, the creation of 3/8 theories should not be prohibited, as it is possible that a non-perfectly mechanical performance of the following music could create such a strong 3/8 feel that listeners would switch their metrical interpretations. For example, the quarter notes in the ascending scale could be stressed very strongly, and the bars following the scale could continue to reinforce the 3/8.

The creation of meter agencies is handled by three separate agencies, one which creates simple beat-levels, one which creates meter agencies by creating new beat-levels as higher levels (inducing higher levels), and one which creates new meter agencies by finding an existing higher level for an existing beat-level (attaching higher levels).

Creating Simple Beat-Levels

The first step in the beat perception process is the creation of simple beat-levels, as all other processing results from their presence. Every time a new note is heard, a check is made to see if the new note and any previously heard note forms a chunk whose length is within minimum and maximum beat intervals mentioned above (.28 to 1.5 seconds for this implementation). If this condition is met, and if there is not already an equivalent beat-level, then a new simple beat-level is created.

Two beat-levels are equivalent if their beat intervals are within the match tolerance mentioned above, and if the last note they identified as a beat is the same. Meter agencies are equivalent if their subdivision structure is the same and if all their beat-levels are equivalent.

The creation process is always active, but is most important at the very beginning of a piece when there are no existing beat-levels. Generally after a few bars, simple beat-levels will have been created for all of the inter-onset intervals within the appropriate range, and unless a significantly different interval appears, no new simple beat-levels will be created.

Inducing Higher Levels

The creation of simple beat-levels will never create meter agencies (i.e. multiple-level beat-levels), nor will it create beat-levels longer than a second or two. Meter agencies are created by an agency which induces a higher level beat-level from the notes which an existing beat-level has identified to be beats.

This agency proceeds as follows: For every beat level that identifies the most recently heard note as a beat, an examination is made of the chunks of time between the most recently heard note and some number of previously identified beats. The size of the chunks in question are limited to be at least two beats long, but no longer than some maximum. This maximum represents the greatest allowable subdivision (for this implementation, the maximum is 5).

If there is a chunk whose interior notes are less important than its delineating notes, and if there is not already an equivalent meter agency, then a new beat-level is created. This new beat-level is connected to a copy of the original beat-level as the higher level of a new top-level meter agency. The original beat-level is copied because the induction of a particular higher level may be an error.

As an example, figure 43 shows a single beat-level which has just identified the most recently heard note as a beat.

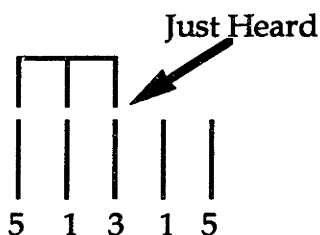


Figure 43.

The chunk of time between the first note and the note just heard is an integral multiple of the beat-level's beat interval. Furthermore, the chunk's

interior note (the second note) is less important than either of its delineating notes (the first and third notes). The following two-level meter agency is thus created:

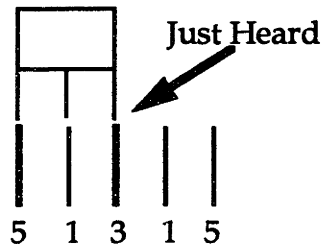


Figure 44.

When this meter agency progresses to the fifth note, another higher level will be induced:

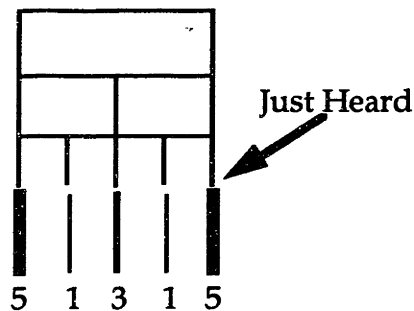


Figure 45.

Thus, higher level beat-levels are induced when there is a suggestion of higher level organization to important notes. The induction creates a new top-level meter agency.

Attaching Higher Levels

The last creation process is noticing that an existing beat-level happens to subdivide the beats of another. Unless there is already an equivalent meter agency, a new meter agency is created from copies of the two original beat-levels. As an example, consider the following four note sequence:

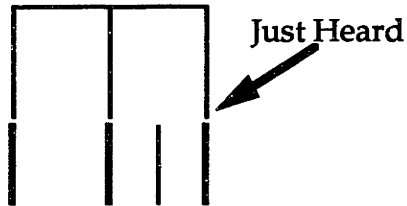


Figure 46.

A beat level which was created from the interval between the first and second notes identifies the fourth note as a beat. The interval between the second and third note would also create a beat level which is shown below:

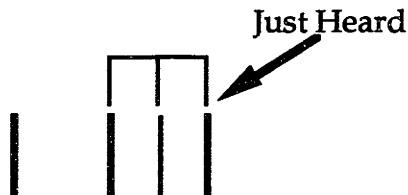


Figure 47.

This beat level subdivides the first, so that the two would be combined into a new meter agency made from copies of both, forming a new top-level meter agency:

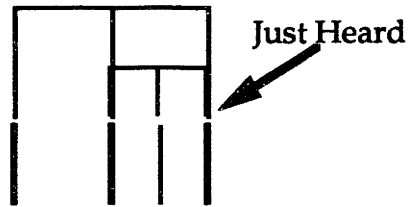


Figure 48.

The creation of meter agencies can be summarized as follows:

1. Simple beat-levels are created from the inter-onset intervals in the music.
2. As these beat-levels identify beats, higher levels are induced from patterns of important notes, creating meter agencies.
3. Meter agencies are also created from beat-levels that “naturally” subdivide.

Managing Meter Agencies

Although the processes outlined above can create many meter agencies, the greatest number of new agencies are created as notes are identified as beats. Recall that every time a meter agency identifies a beat, it replicates itself, representing one agency that identified the beat and one that did not. As a normal case, meter agencies will be created and replicated at virtually every note. Unless limited, the number of meter agencies in the system will grow exponentially over time, although as previously discussed, a perceptual model cannot allow unbounded growth. Clearly, some method must be implemented for managing and limiting the profusion of agencies.

The top-level meter agencies which are created by replication as a product of beat identification form a logical association. They collectively represent the feasible assignments of beats to notes of a specific metrical structure and beat value, starting from a particular note. These agencies are thus grouped together as subordinates in a managing agency called a *theory*.

The limiting of meter-agencies occurs at two levels. There is a limit imposed on both the number of theories which can exist, and on the number of meter agencies within an existing theory. Meter agencies are limited by keeping only a certain number of the highest scoring meter agencies in a theory. Theories are limited the same way, using the score of the highest scoring meter agency.

Implementation

This section describes the implementation of the above model as a computer program. The implementation was coded in Macintosh Allegro Common Lisp (Apple Computers) and runs on a Macintosh II computer. Extensive use was made of the Objectlisp (Drescher 1987) object oriented programming extensions to Common Lisp. Additionally, the implementation relied on tools provided by Hyperlisp (Chung 1988), a real-time MIDI programming environment developed at the Massachusetts Institute of Technology Media Laboratory in conjunction with the Hyperinstruments research project (Machover & Chung 1988).³⁴

The implementation of this model is surprisingly compact. The source code, which comprises less than ten pages, is included as an appendix.

³⁴ Hyperlisp is available from the author for non-commercial use.

Input

Input to the system originates either from MIDI data recorded directly into Hyperlisp, or from recordings made on a Bosendorfer 290 SE recording piano which were converted to the Hyperlisp internal MIDI format. Input is then converted from Hyperlisp's format to an *event-list* which retains only the note onset and offset data.

Note Importance

As mentioned before, the note importance agency is effectively ignored in this model. Note duration is used as a crude measure of importance.

Objects and Agencies

Each agent described above was implemented as an Objectlisp object. Agencies are represented as objects which "point" (or refer) to subordinate agents. For example, the simple beat-level agent is implemented as a single Objectlisp object. A meter agency is just a beat-level object that points to a lower level beat-level.

Discussion

Rhythmic and Metric Interpretation

An interesting property of the model presented above is the manner in which rhythmic and metric interpretation are handled. Earlier discussion concluded that rhythmic and metric interpretation proceed simultaneously, but that a correct rhythmic interpretation was not strictly necessary for a correct metric

one. The important thing is that the rhythmic and metric interpretations be compatible.

While the meter agencies presented above do not specifically attempt to produce a rhythmic interpretation, they do identify the compatibility constraints. Consider the following example which shows the notes a beat-level agent has identified to be beats:



Figure 49.

Whereas the exact rhythmic interpretation of every note is not known, the beat-level has imposed certain constraints. The first three notes, for example, must fit into the time of one beat, although the intended rhythmic values are ambiguous. If, however, this beat-level is the higher level in the meter agency depicted below, the rhythmic interpretation is further constrained.



Figure 50.

According to this meter agency, the first note's rhythmic value must be half a beat. The values of the second and third notes are still unknown, but they are now constrained to fit in the time of half a beat.

Although the higher level in the example above will achieve a higher score if its lower level performs well, it is tolerant to errors in its lower level. If, for example, the intended rhythmic interpretation of the first three notes above happened to be eighth-note triplets, the erroneous identification of the second note as an eighth-note by the lower level does not adversely affect the higher level. It is unnecessary to entirely resolve rhythmic ambiguity.

Ambiguity

Ambiguity is modeled well in the proposed model. In addition to pursuing several different interpretations, the beat perception agency is capable of determining the degree of ambiguity. The interpretations are available as the highest scoring theories, and the degree of ambiguity can be determined by the distribution of the scores. In the case where a single theory has a much higher score than all of the others, there is little ambiguity. On the other hand, if all of the highest scoring theories have roughly the same score, the meter is ambiguous.

The ability to determine the degree of ambiguity is important to the interactions between a beat perception agency and other music perception agencies. In cases where the meter is intended to be ambiguous, the beat perception agency is capable of detecting this condition and providing the competing theories. The model can qualify its results with a measure of certainty.

A "Society of Mind" Model

One of the most significant aspects of the model is that it demonstrates the tractability of the beat perception problem to a "Society of Mind" approach. The intelligence of the model is derived from the hierarchy of connected agents. There is no centralized control mechanism which makes the critical decisions. Furthermore, the individual agents interact only with a small number of superiors and subordinates. Each agent performs simple tasks and makes simple decisions based on a limited domain of input. On the other hand, no attempt has been made to reduce the problem into a single underlying principle that is intended to explain the entirety of beat perception. While the concept of matching adjacent chunks of time is fundamental to the model, the management of the matching process is equally important.

Additionally, the organization of the model's agents fit well into the "X-Brain" structure previously discussed. Figure 51 below depicts the organization of the model into three "brains:"

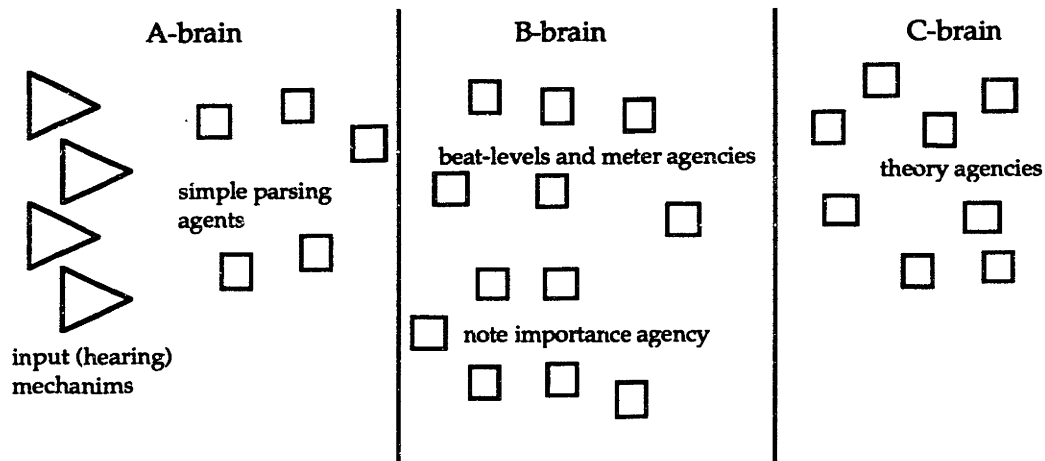


Figure 51.

Each “brain” acts only on the preceding brain’s agents. The A-brain is responsible for simple parsing of the raw acoustic signal into features such as note onsets and offsets. The B-brain examines only the behavior of the parsing agents and proceeds with assignment of note importance and the matching of time chunks via beat-level agents. The C-brain is responsible for managing the beat-levels via theory agencies. It has no knowledge or interest of the A-brain.

Problems

As previously mentioned, the model presented above allows for significant deviation of inter-onset intervals, but does not track tempo changes. This is not a difficult limitation to overcome. Since beat-levels can detect when the beats identified are early or late, the beat intervals can be adjusted accordingly. On the other hand, it is likely that tempo tracking should occur at a higher level so that there is better coordination between connected beat-levels. For example, one would like acceleration within the measure to create the

expectation of an earlier down-beat. One way of handling this is to create a separate tempo tracking agency which monitors the performance of the highest scoring theory. When the tempo tracking agency detects acceleration, the beat intervals at all beat-levels are adjusted for the entire theory.

Additionally, the model does not adequately handle meter changes. Although the model would eventually discover the new meter for a single change, there is no mechanism for expecting repeated changes. For example, the model would fail on a piece whose metrical structure consists of repetitions of two 4/4 measures followed by a 3/4 measure. Extreme tempo changes present another problematic case. The scheme for tracking tempo mentioned above would probably fail when changes occur too rapidly as in extreme *accelerando* or *rallentando*. Tempo tracking would only work for “normal” cases where the tempo drifts relatively slowly.

Both of the above problems can be alleviated through the use of higher level agencies which can examine and modify the beat perception agency, as in the “X-brain” model previously discussed. In the case of regular meter changes, a higher level agency can recognize the pattern of changes, and inform the beat perception agency when such a change is likely to take place. For extreme tempo changes, an agency which can take a larger view of musical phrases can detect when an *accelerando* or *rallentando* is taking place, and again inform the beat perception agency.

A further problem with the model is its inefficiency. There is a natural abundance of redundant computation. For example, when a beat-level induces a higher level, both the original beat-level and its copy, which is now

part of the new meter agency, will exist in the system. For the most part, the original and the copy will independently discover the same beats, and hence, will perform the redundant computations.

It should be noted that an inefficient model may accurately reflect human beat perception. There is no way of knowing how inefficient our own perceptual processes are. However, if the goal is to implement a practical beat perception system to assist in automatic transcription, for example, these inefficiencies can be costly.

Eliminating the inefficiencies is not a trivial problem. The power of the system is derived, in part, from the ability to pursue many parallel theories simultaneously without needing to coordinate between them. The inefficiency is the price paid for a simple management mechanism. Eliminating redundant computation implies sharing the processing of individual beat-levels in a fairly sophisticated way. This sharing may violate the constraints of a "Society of Mind" model.

Conclusions

The work in this thesis has attempted to show that musical problems must be treated as cognitive processes in order to successfully model important underlying interactions. A cognitive perspective on beat perception allows computational access to phenomena such as expectation and ambiguity that are inaccessible to syntactic approaches.

While the model and implementation presented above is far from a robust, practical system, it successfully addresses a number of important beat perception issues that others have ignored. For the sake of simplicity, many important aspects of beat perception have been neglected (such as note importance, tempo tracking, and theory persistence), but the framework established by this model is designed to accommodate these aspects. Instead of focusing exclusively on one idealized sub-problem of beat-perception, as many researchers have done (such as Longuet-Higgins and Lee (1984), Steedman (1977), and Lee (1985)), the model presented here is intended to be extendable to encompass the phenomena of beat perception as it is currently understood.

Although any research topic requires focus, the findings which result from the solutions of over-idealized problems are of little use unless the solution can be incorporated into a more general system. The establishment of an extendable framework for music perception is an obvious goal. The *Society of Mind* model is especially attractive in this context, as it represents a global perspective on the human mind that is implementable on a computer. The

results of this thesis demonstrate the tractability of at least one aspect of music perception to a cognitive approach.

Future Research Directions

In addition to the tempo tracking and meter changing problems mentioned above, there are a number of other research directions indicated by the work in this thesis.

Note Importance Agency

The most significant omission from this thesis is a thorough model of the note importance agency. This complex agency needs to integrate data from several sources such as stress, agogic accent, melodic repetition, expressive timing, etc., into a single dimension of importance. As mentioned earlier, this is a circular process, requiring the output of the beat perception agency to compute all of the contributions. The exploration of this agency will require exactly the general framework for combining many different processes discussed above.

Non-Monophonic Music

The extension of this system, or any monophonic system, into a non-monophonic environment is a very difficult undertaking. One scheme might be to apply the above model to each of the voices separately, and then take intersection of high scoring theories as the overall interpretation. This approach may ignore important rhythmic information that is only apparent when all voices are considered at once (of course, the model could be applied once more to all the voices). The key to the problem is understanding how

multiple voices rhythmically interact, but this is an extremely involved question.

Live Beat Tracking System

Although the beat perception agency embodies a real time algorithm, it is not a live beat tracking system like Dannenberg and Mont-Reynaud's (1987). It requires a bounded, but lengthy time to compute its response to each note (typically one or two seconds). A future research direction to be pursued is simplifying and optimizing the system for live use. The applications of such a live system are discussed below.

Possible Applications

Automatic Transcription

An obvious application of this thesis is in automatic transcription. Although the beat perception agency does not explicitly recover note values, it provides enough constraints on rhythmic interpretation that producing a notation from the metric interpretations should be a relatively easy task. Because the agency also maintains multiple interpretations, a human overseer can select from the feasible possibilities in ambiguous cases. In the automatic transcription context, the model might best serve as a "notationist's assistant."

Hyperinstruments

The original motivating force behind the work in this thesis has been the desire to create a real time system for live performance as a part of a Hyperinstrument (Machover & Chung, 1988). Hyperinstruments are

interactive, virtual instruments which intelligently map performance gesture to a sophisticated result. A real time beat perception system would enable the hyperinstrument to understand a great deal about the rhythmic component of the music, allowing synchronization between live playing and prerecorded events, or between several players. Additionally, a hyperinstrument system which is aware of the beats and meter could use that information to drive other multi-media outputs such as video images, computer animation, lights and set movements. The integration of musically intelligent systems into live performance portends a new and exciting age of musical performance.

Appendices

Hyperlisp

Hyperlisp is a real-time, MIDI programming environment embedded in Macintosh Allegro Common Lisp. It is designed for facilitating the development of real time music systems which use MIDI. The system was developed specifically for the Hyperinstruments project at the MIT Media Laboratory, and is optimized for interactive systems which require fast response times. Hyperlisp provides two main services for the music programmer: routines for midi processing and primitives for scheduling the application of functions in real-time. Programs written in Allegro Common Lisp can use these services for a wide variety of real-time MIDI systems.

The Hyperlisp system is based on a simple non-prioritized scheduler which essentially provides two facilities for hyperinstrument programmers: delayed function application and notification of the arrival of MIDI input. Hyperlisp also provides various routines for processing MIDI data, and is linked in with ObjectLisp, the object oriented programming extension provided by Allegro Common Lisp. A simple MIDI delay program written in Hyperlisp is shown in below:

```
(defobject midi-delay midi-object)           ; define a midi-delay object

(defobfun (exist midi-delay) (args)         ; called on object creation
  (have 'delay 50)                          ; delay in 1/100's second
  (add-midievent-handler self midi-in))     ; cause midi-in to be called
                                           ; whenever there is MIDI
                                           ; input

(defobfun (midi-in midi-delay) (event)      ; called on MIDI input
  (out event)                               ; write event now
  (post delay self (out event)))           ; and again after the right
```

```

; delay
(defobjfun (out midi-delay) (event)
  (midi-write event)) ; write out event

```

An Example Hyperlisp Program

When a `midi-delay` object is created, its `exist` function is called, and `add-midievent-handler` adds the object's `midi-in` function to the functions Hyperlisp should call whenever there is MIDI input. A `midievent` handler always takes one argument, the `midievent`. A `midievent` is represented in Hyperlisp by a thirty-two bit integer whose lower twenty four bits are the MIDI status byte and the two data bytes.

Once the object is created, Hyperlisp will call `midi-in` for every `midievent` it receives from the MIDI controller. `Midi-in` does two things: it calls `out` which forwards the event to the synthesizer immediately, and it schedules `out` to be applied to the same event `delay` centiseconds into the future. When that amount of time has passed, regardless of what else may be going on in the system, the scheduler will apply `out` to the proper event. The result is a MIDI delay analogous to a digital or analog delay in a conventional sound effects processor.

The basic Hyperlisp algorithm is as follows: Wait for either MIDI input or a clock tick. If there is MIDI input, call every function that requested notification via `add-midievent-handler`. If there is a clock tick, check if there are any function applications that were scheduled via `post` for the current tick. If so, apply those functions one by one. Arguments to a delayed function are evaluated when the function is scheduled in the scheduling

object's environment. All times are expressed in 1/100's of a second (centiseconds).

Bibliography

- Bamberger, J. (1980). Cognitive Structuring in the Apprehension and Description of Simple Rhythms. *Archives de Psychologie*, **48**, 171-199.
- Benjamin, WE. (1984). A Theory of Musical Meter. *Music Perception*, **1**(4), 355-413.
- Bengtsson, I. & Gabrielsson, A. (1980). Methods for Analyzing Performances of Musical Rhythm. *Scandinavian Journal of Psychology*, **21**, 257-268.
- Bengtsson, I. & Gabrielsson, A. (1983). Analysis and Synthesis of Musical Rhythm. In J. Sundberg (Ed.), *Studies of Music Performance, Publications issued by the Royal Swedish Academy of Music*, **39**, 27-60.
- Chowning, J., Rush, N., Chafe, C., Schloss, WA., & Smith, J. (1984) Intelligent Systems for the Analysis of Digitized Acoustical Signals. Stanford University CCRMA Report, STAN-M-15.
- Clarke, EF. (1987). Categorical Rhythm Perception: An Ecological Perspective. In Gabrielsson (Ed.), *Action and Perception in Rhythm and Music*. Stockholm: The Royal Swedish Academy of Music, 19-33.
- Cooper, G. & Meyer, L. (1960). *The Rhythmic Structure of Music*. Chicago: The University of Chicago Press.
- Dannenberg, RB., & Mont-Reynaud, B. (1987). Following an Improvisation in Real Time. *Proceedings of the International Computer Music Conference*.
- Dannenberg, RB. (1988). Personal Correspondance.
- Desain, P. & Honing, H. (1988). Quantization of Musical Time: A Connectionist Approach. Unpublished Manuscript.
- Drescher, G. Object Oriented Logo. (1987). In Mahler and Yazbani (Eds.), *Artificial Intelligence and Education*. Norwood, N.J.: Ablex Publishers.
- Essens, P., & Povel, DJ. (1985). Metrical and Nonmetrical Representations of Temporal Patterns. *Perception & Psychophysics*, **37**(1), 1-7.
- Fraisse, P. (1987). A Historical Approach to Rhythm as Perception. In Gabrielsson (Ed.), *Action and Perception in Rhythm and Music*. Stockholm: The Royal Swedish Academy of Music, 7-18.

Fraisse, P. Rhythm and Tempo. (1982). In Deutsch (Ed.), *The Psychology of Music*. New York: Academic Press, 149-180.

Gabrielsson, A. (1973). Similarity Ratings and Dimension Analyses of Auditory Rhythm Patterns. I and II. *Scandinavian Journal of Psychology*, **14**, 138-176.

Gabrielsson, A. (1985). Interplay Between Analysis and Synthesis. *Music Perception*, **3**(1), 59-86.

Gabrielsson, A. (1988). Timing in Music Performance and Its Relations to Music Experience. In Sloboda (Ed.), *Generative Processes in Music*, Oxford: Clarendon Press.

Gabrielsson, A., Bengtsson, I., and Gabrielsson, B. (1983). Performance of Musical Rhythm in 3/4 and 6/8 Meter. *Scandinavian Journal of Psychology*, **24**, 193-213.

Handel, S. & Oshinsky, J. (1981). The Meter of Syncopated Auditory Polyrhythms. *Perception & Psychophysics*, **30**(1), 1-9.

Handel, S. & Todd, P. (1981). Segmentation of Sequential Patterns. *Journal of Experimental Psychology: Human Perception & Performance*, **7**(1), 41-55.

Hirsh I. & Sherrick C. (1961). Perceived Order In Different Sense Modalities. *Journal of Experimental Psychology*, **62**(5), 423-432.

Howell, P, Cross, I, West, R. (1985). (Eds.), *Musical Structure and Cognition*. London: Academic Press.

Jones, MR. (1976). Time, Our Lost Dimension: Toward a New Theory of Perception, Attention, and Memory. *Psychological Review*, **83**(5), 323-355.

Lee, CS. (1985). The Rhythmic Interpretation of Simple Musical Sequences. In P. Howell, I. Cross, & R. West (Eds.), *Musical Structure and Cognition*. London: Academic Press.

Lerdahl, F. & Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. Cambridge: MIT Press.

Lewis H. & Papadimitriou, C. (1981). *Elements of the Theory of Computation*. New Jersey: Prentice-Hall.

Longuet-Higgins, HC & Lee, CS. (1982). The Perception of Musical Rhythms. *Perception*, **11**, 115-128.

Longuet-Higgins, HC & Lee, CS. (1984). The Rhythmic Interpretation of Monophonic Music. *Music Perception*, 1(4), 424-441.

Machover, T., & Chung, J. (1988). Hyperinstruments: Musically Intelligent/Interactive Performance and Creativity Systems. Massachusetts Institute of Technology Media Laboratory Internal Memo.

Minsky, M. (1986). *The Society of Mind*. New York: Simon and Schuster.

Minsky, M. (1981). Music, Mind, and Meaning. *Computer Music Journal*, 5, 28-44.

Monahan C. & Carterette EC. (1985)., Pitch and Duration as Determinants of Musical Space. *Music Perception*, 3(1), 1-32.

Mont-Reynaud, B. & Goldstein, M. (1985). On Finding Rhythmic Patterns in Musical Lines. *1985 Proceedings of the International Computer Music Conference*.

Musical Instrument Digital Interface. Specification 1.0, 1986.

Palmer, C. & Krumhansl, C. (1987a). Pitch and Temporal Contributions to Musical Phrase Perception: Effects of Harmony, Performance Timing, and Familiarity. *Perception & Psychophysics*, 41(6), 505-518.

Palmer, C. & Krumhansl, C. (1987b). Independent Temporal and Pitch Structures in Determination of Musical Phrases. *Experimental Psychology: Human Perception & Performance*, 13(1), 116-126.

Povel, DJ. (1981). Internal Representation of Simple Temporal Patterns. *Journal of Experimental Psychology: Human Perception & Performance*, 7, 3-18.

Povel, DJ., & Essens, P. (1985). Perception of Temporal Patterns. *Music Perception*, 2(4), 411-440.

Rosenthal, D. (1989). A Model of the Process of Listening to Simple Rhythms. *Music Perception*, 6(3), 315-328.

Rowe, R. (1989). Short-term Memory and the Perception of Pulse. Massachusetts Institute of Technology Media Laboratory Internal Memo.

Sachs, C. (1953). *Rhythm and Tempo, A Study in Music History*. New York: Columbia University Press.

Schloss, WA. (1985). On the automatic Transcription of Percussive Music -- from acoustic signal to high-level analysis. Stanford University CCRMA Report, STAN-M-27.

Serafine, ML. (1988) *Music As Cognition*. New York: Columbia University Press.

Shaffer, LH, Clarke E., & Todd, NP. (1985). Metre and Rhythm in Piano Playing. *Cognition*, 20, 61-77.

Shaffer, LH & Todd, NP. (1987). The Interpretive Component In Musical Performance. In Gabrielsson (Ed.), *Action and Perception in Rhythm and Music*. Stockholm: The Royal Swedish Academy of Music, 139-152.

Sloboda, JA. (1985). *The Musical Mind*. Oxford: Clarendon Press.

Steedman, Mark. (1977). The Perception of Musical Rhythm and Metre, *Perception* 6, 555-569.

Todd, NP. (1985). A model of expressive timing in tonal music. *Music Perception*, 3(1), 33-59.