# Rhythmic Complexity in Jazz: An Information Theory Approach

Rhythmic Complexity in Jazz:
An Information Theory Approach

A Dissertation Presented

by

Douglas R. Abrams

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

February 2023

Department of Music and Dance

Rhythmic Complexity in Jazz:
An Information Theory Approach

A Dissertation Presented

By

Douglas R. Abrams

Approved as to style and content by:

_____

Christopher White, chair

_____

Brent Auerbach, member

_____

Jason Yust, member

_____

Matthew Westgate, DEPARTMENT CHAIR

Department of Music and Dance

DEDICATION

To my parents

## ACKNOWLEDGEMENTS

ABSTRACT

RHYTHMIC COMPLEXITY IN JAZZ:

AN INFORMATION THEORY APPROACH

FEBRUARY 2023

DOUGLAS R. ABRAMS, S.B. MASSACHUSETTS INSTITUTE OF

TECHNOLOGY

M.M. MANHATTAN SCHOOL OF MUSIC

M.M. UNIVERSITY OF MASSACHUSETTS, AMHERST

Ph. D. UNIVERSITY OF MASSACHUSETTS, AMHERST

Directed by: Christopher W. White

Many techniques of quantifying rhythmic complexity have been explored, including methods based on the concept of entropy. Roughly speaking, entropy measures a rhythm's unpredictability. The primary goals of this study were to answer two questions: 1) Does rhythmic entropy correspond to perceived rhythmic complexity? and 2) Does entropy of a jazz solo depend on soloist? Additionally, I used entropy to study the relationship between sheet music and jazz versions of songs from the American songbook, and I used the concept of mutual information to study soloist-accompanist interactions in the music of Charlie Parker.

I asked fifteen UMass music majors to rate short, eighth-note based jazz rhythms for complexity. Entropies were calculated by constructing distributions based on the inter-onset intervals (IOI's) between notes. Using a mixed effects multiple regression model, I found, as expected, that higher entropy resulted in higher complexity ratings. Other factors did, too, namely: number of notes, syncopation, lack of periodicity, and the effects of each complexity rating on the following one. It is possible that entropy was mediated by lack of periodicity.

I then transcribed (or compiled and checked) a corpus of 88 solos by Armstrong, Hawkins, Young, Christian, and Parker, and calculated entropies based on the IOI's between stress-accented notes. I used the technique of estimated marginal means with number of distinct IOI's and number of accents as covariates to show that entropy depends significantly on soloist: solos by Lester Young were significantly more entropic

than those by Armstrong, Christian, and Parker. Stress accent density and contour accent density were used to explain the unexpected lack of differentiation between Parker and Hawkins in terms of entropy.

I demonstrated that jazz renditions of popular songs had higher entropy than their sheet music counterparts. Finally, I used mutual information to show that interrelationships between Parker and his accompanists were stronger than those between Parker and a Charleston comping rhythm.

This work demonstrates the utility of entropy-based methods in predicting a listener's perceived complexity, in characterizing a soloist's oeuvre, and in describing embellished versions of songs. It also demonstrates the utility of mutual information in describing soloist/accompanist interactions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

# CHAPTER 1

# INTRODUCTION

The field of research on rhythmic complexity is a fertile one. In recent decades, a plethora of methods for quantifying rhythmic complexity has been introduced. Nonetheless, an exact definition of rhythmic complexity is hard to come by. One intuitive definition might be that rhythmic complexity reflects the difficulty of performing a given rhythm. Another might reflect the difficulty of tapping along with a rhythm, or recalling a rhythm after a certain amount of time has elapsed. And in some cases, a listener might *imagine* how difficult it would be to perform a particular rhythm. In fact, each of these definitions has been used, e.g. in studies by Shmulevich and Povel (2000), and Fitch and Rosenfeld (2007). In any case, the approach taken here asks experimental subjects and readers to rely on their own intuitive ideas about rhythmic complexity in order to provide experimental data or interpret the current work's results.

The study of rhythmic complexity allows us to compare complexity in music to complexity in other fields. This is partially due to the fact that at its most basic, music can be understood as a form of communication. Semiotically speaking, of course, it is more vague than other forms of communication such as speech – the relationships between signifier and signified are less specific in music than in speech – but it is useful nonetheless to treat music as communication, for reasons discussed below. Though in most cases it is unlikely that any two different listeners will agree 100% on what a given piece of music communicates, they might agree more often on how it communicates – and the analysis of how a piece communicates is the realm of music theory.

Beyond speech and music, of course, there are many other forms of communication as well, notably the digital transmission of sound or video; the theory undergirding such types of communication dates to the mid-20th century, and goes under the name of *information theory*. Shannon (1948) laid the groundwork for this field in his seminal work *A Mathematical Theory of Communication*. Since then, information theory has been applied in many fields outside of communication theory, including music.

How is it possible that a theory devised for the electrical transmission of signals can also be applied to music? The answer comes from Leonard Meyer and his work linking expectation in music – specifically, *thwarted* expectation – to meaning (1956, 1957). Meyer posits that low probability events in the context of a given piece (that is, events which thwart our expectations) have more information, and thus more meaning, than high probability events. Keeping in mind that, quoting Joel Cohen (1962), "statistical probability, or relative frequency, corresponds to the listener's expectations," there is then a natural correspondence between information theory and meaning in music. To wit, information theory deals with Cohen's probabilities – thus, with expectation – and these comprise the *information content* of a set of events; hence the correspondence between music theory and information theory.

Thus, treating music as a form of communication allows us to apply the techniques of information theory to music. That is what is done here.

Shannon Entropy (or just "entropy"), introduced in *A Mathematical Theory of Communication*, is information theory's most fundamental concept, and has been employed in many fields. Roughly speaking, it measures the "unpredictability" of a sequence of events. Its wide applicability is probably due to its simplicity: as will be seen

in the next chapter, any quantity that can be described by a probability distribution (defined later) can be described, using a simple formula, by entropy. As a quick Wikipedia search will indicate, fields in which entropy has been applied include statistical inference, cryptography, neurobiology, perception, bioinformatics, thermal physics, quantum computing, information retrieval, intelligence gathering, plagiarism detection, pattern recognition, anomaly detection, and others. Therefore, it should not come as a surprise that it has been applied in the field of music theory.

A cluster of papers applying information theory to music appeared in the fifties and early sixties (Pinkerton 1956; Meyer 1956, 1957; Youngblood 1958; Cohen 1962), applying the concept of entropy to pitch structures and laying the philosophical groundwork for the field. A PhD dissertation in 1959 (Crawley) pioneered the application of information theory to rhythm, and a single paper applying entropy to both pitch and rhythm (Hiller and Fuller) appeared in 1967. Several papers in the eighties and early nineties refined the earlier work (Knopoff and Hutchinson 1981; Knopoff and Hutchinson 1983; Snyder 1990), and since 2002, there have been a number of papers and books on information theory and music, going beyond entropy-based analysis and working with the information content of individual events and with a generalization of entropy called Markov chains (presaged by Hiller and Fuller 1967): Abdallah 2002; Pearce and Wiggins 2004; Sadakata et al. 2006;.Huron 2006; Temperley 2007; and Margulis and Beatty 2008. Finally, since the late nineties and early two-thousands, a number of papers using Markov chains to analyze and compose jazz have appeared, including: Yarom (1997); Gillick, Tang, and Keller (2010); Pfeiderer and Krizler (2010); Bell (2011); Victor (2012); Norgaard, Spencer, and Montiel (2013); Linskens and

Schoenmakers (2014); Yun (2016); Rouse (2017); Quick and Thomas (2019); and Frieler (2019).

To my knowledge, only three previous studies have examined whether or not the concept of entropy can be used to describe the effects the perception of rhythmic complexity: Toussaint and Thul (2008), de Fleurian et al. (2014) and de Fleurian et al. (2017). These are summarized in Chapters 3 and 4. Furthermore, while a large number of studies applying multiple-order Markov chains to jazz have appeared in the last two decades, fewer studies have worked explicitly with zeroth-order chains, that is, with entropy. The present work addresses these lacunae, and serves two secondary purposes as well.

Primarily, this work provides preliminary evidence that a particular implementation of the concept of entropy can be used to describe the perception of rhythmic complexity, and shows that, in a corpus of 88 transcribed solos by five great jazz musicians, entropy depends to a certain extent on soloist. Secondarily, this work uses entropy to describe the phenomenon of melodic embellishment, and uses the concept of mutual information to study the interaction between Charlie Parker and his accompanists in a limited corpus of 10 transcribed solos.

The chapter immediately following this one (Chapter 2) introduces entropy formally, as well as the related topics of normalized entropy, redundancy, cross entropy, relative entropy, mutual information, Markov chains, and surprisal. Entropy is shown to be the uncertainty of a "random variable" (defined in Chapter 2), and normalized entropy and redundancy reflect this in modified form. Cross-entropy is introduced as a means of evaluating a model's fit to a given set of data. Relative entropy is introduced as a kind of

"distance" between two random variables, and mutual information is shown to be how "far away" two variables are from being completely independent. Markov chains are introduced as a way of studying the probability of events depending on one or more previous events. Finally, surprisal is introduced as the contribution to the entropy of a single event.

The following chapter (Chapter 3) is a review of literature, situating the current research in the context of pre-existing work. It covers not only information theory and music, but also probability theory and music, rhythmic complexity, Markov chains and jazz improvisation, and jazz history and the evolution of rhythm.

Chapter 4 covers experimental methodology and results. In it I demonstrate that for rhythms that are multiples of eighth-notes and eighth-rests, an implementation of entropy involving rhythm, periodicity or the lack thereof, syncopation, and the number of accents (for a fixed excerpt length) are significant factors in the perception of rhythmic complexity in jazz.

Chapter 5 covers computational methodology and results. With a corpus of 88 solo transcriptions, I use a technique that takes two covariates into account to show that the implementation of entropy alluded to above depends to a certain extent on soloist amongst a group of five soloists selected for this study: Louis Armstrong, Coleman Hawkins, Lester Young, Charlie Christian, and Charlie Parker.

Chapter 6 contains close readings of eleven excerpts: four pairs grouped according to values of the two covariates introduced in Chapter 5, one pair comprised of readings of *Body and Soul* by Hawkins and Young, and one excerpt by Young selected because it has the highest entropy in the corpus.

Chapter 7 covers melodic elaboration and entropy, and demonstrates that jazz renditions of popular songs have substantially higher entropies than sheet music versions.

Chapter 8 shows how mutual information can be used to study soloist/accompanist interactions. This approach reveals that mutual information for Parker and his accompanists is greater than that for Parker and a simple comping rhythm and less than that for Parker and a random comping rhythm, and explains an outlier by referring to the presence of a less experienced pianist on the corresponding recording.

# CHAPTER 2

# A BRIEF INTRODUCTION TO INFORMATION THEORY

## 2.1 Introduction

A "random variable" is a variable which, when measured, can take on one of

several values with a given probability of occurrence for each value[1]. The set of values a

random variable can take on is called its "sample space," and the set of probabilities with

which a random variable takes on its various values is described by its "probability

distribution." The probability with which a given value of the random variable occurs

ranges from zero (no probability of occurring; this value would not technically be a part

of the probability distribution) to one (100% probability of occurring; this value would be

the only value in the probability distribution). The sum of all values represented by a

probability distribution is always one, since *something* must occur.

Consider two simple examples: first, the case of a fair six-sided die. Define a

random variable "Fair" with "1", "2", "3", "4", "5", and "6" as its sample space.

Measuring this random variable would mean tossing the die and recording the outcome.

Since each outcome is equally likely, this variable's probability distribution would be as

shown in Table 1.

| Value | 1 | 2 | 3 | 4 | 5 | 6 |
|-------------|-----|-----|-----|-----|-----|-----|
| Probability | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 |

**Table 1.** Probability distribution for random variable "Fair"

Note that the probabilities sum to one.

---

[1] I will focus here on "discrete" random variables, since the calculated entropy values I use are constructed from discrete random variables. This is not affected by the fact that the entropy values themselves are treated as continuous variables.

Now consider a loaded die. Define a random variable "Loaded." Its sample space would be the same as that for the variable "Fair," but the probability distribution would be different. It might look something like Table 2.

| Value | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Probability | 1/5 | 1/5 | 1/5 | 1/5 | 1/10 | 1/10 |

**Table 2.** Probability distribution for random variable "Loaded"

Note that the probabilities still sum to one.

In our case, the random variables we will be working with measure the inter-onset intervals (IOIs) between dynamically accented notes in solos by well-known jazz musicians. Roughly speaking, dynamically accented notes are simply notes which are louder than surrounding notes (Roeder 2017); a more detailed explanation of how accented notes are identified is given in Chapter 5. Inter-onset intervals are defined as the durational distances between the attacks or onsets of pairs of notes (Patel and Danielle 2003). While the time delays between accented notes in a solo are not, of course, actually *random*, we can glean important information about a solo (in particular, about the unpredictability of its rhythms) by treating it as though time delays have been selected at random from the distribution of time delays describing that solo.

Figure 1 provides an example of how I obtain probability distributions from solos; there are other ways, of course, but this is the method I use. Figure 1 is a brief excerpt from Lester Young's solo on "Blues for Greasy;" accent marks indicate dynamic accents. Having identified dynamically accented notes, the next step is to tabulate the inter-onset intervals.

**Figure 1.** Musical example for calculating a probability distribution

Note that for solos in which the eighth-notes are "swung," care must be taken to identify when an accent falls on the long or short part of the beat, since this will affect the IOIs involving that accent. In this example, there are two examples of IOIs that depend on whether they start with long eighths or short eighths: seven eighth-notes starting on the long part of the beat (beat one in m. 2 to the "and" of four in m. 2; beat one in m. 5 to the "and" of four in m. 5), and seven eighth-notes starting on the short part of the beat (the "and" of one in m. 4 to beat one in m. 5). A distinction must be made since one of them will have duration L+S+L+S+L+S+L and the other will have duration S+L+S+L+S+L+S, where L is the duration of the long half and S is the duration of the short half of the swung eighth-note pair. (The distinction need not be made for IOIs consisting of an even number of eighth-notes).

We can assign numerical values to each of these IOIs, assuming that the ratio of the long part of the beat to the short part of the beat is 2:1,, in other words, that the length of the long part of the beat is 1.333 eighth-notes and the length of the short part of the beat is 0.666. Note that the ratio is 2:1 and that they sum to two eighth-notes, or one quarter. The IOI of seven eighth-notes starting on the long part of the beat is 7.333 eighth-notes long, while the IOI of seven eighth-notes starting on the short part of the beat is 6.666 eighth-notes long.

With this in mind, the frequency of each IOI may be tabulated as follows: two eighth-notes, one; 7.333 eighth-notes, two; 6.666 eighth-notes, one; ten eighth-notes, one. Note that there are six accents, thus five IOIs. Probabilities are obtained by dividing frequencies by five. The resulting distribution is given in Table 3.

| IOI (eighth-notes) | 2 | 6.66 | 7.33 | 10 |
|---|---|---|---|---|
| **Probability** | 1/5 | 1/5 | 2/5 | 1/5 |

**Table 3.** Probability distribution for excerpt in Figure 1

## 2.2 Entropy

The fundamental quantity used in information theory is called "entropy." Roughly speaking, entropy measures the "uncertainty" of a random variable (Cover and Thomas, 2006). In his 1948 paper, *A Mathematical Theory of Communication*, Claude Shannon derived a formula for entropy based on three axioms, stated below. In what follows, $H$ will be a candidate function for entropy, defined on the probability distribution given by the set of all $p_i$ for $i = 1...n$. These three axioms are the axioms from Shannon's paper:

1. $H$ should be continuous in the $p_i$

2. If all the $p_i$ are equal ($p_i = 1/n$), then $H$ should be a monotonic increasing function of $n$. With equally likely events there is more choice, or uncertainty, when there are more events.

3. If a choice be broken down into two successive choices, the original $H$ should be the weighted sum of the individual values of $H$ ...

*Shannon* 1948, 10

The third axiom deserves more attention. Consider the tree diagram in Figure 2, from

Shannon (1948). On the left we see a diagram with three choices, moving left to right. On

the right we see a diagram with two choices followed by a further two choices if the

lower branch is taken. Following the stipulation in axiom 3, the entropy calculated for the

left diagram would be H(1/2,1/3,1/6) while the entropy for the right diagram would be

H(1/2,1/2) for the first choice, plus (½)H(2/3,1/3) for the second choice. Axiom 3

requires that these two ways of calculating $H$ yield equivalent results.

<figure>
Left diagram: branches labeled 1/2, 1/3, 1/6.
Right diagram: first choice branches labeled 1/2 (upper, to 1/2) and 1/2 (lower); lower choice branches 2/3 (to 1/3) and 1/3 (to 1/6).
</figure>
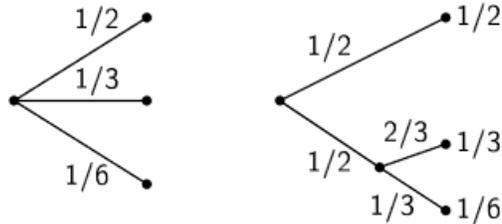
**Figure 2.** Two ways of calculating the same entropy (see third axiom above)

These three axioms are sufficient to derive the following formula for entropy,

modulo a positive multiplicative constant:

$$H \equiv - \sum_x p(x)\log(p(x)) \qquad \text{Eq. 2.1}$$

where *p(x)* is the probability with which a variable $X$ takes on the value $x$ and the sum

(indicated by the Greek letter sigma) is over all possible values of x. The convention

adopted here, common in information theory, will be to use logarithms of base 2, so that

entropy will be measured in what information theorists call "bits."

To confirm that the third axiom has been satisfied, I will apply Eq. 2.1 to the

example illustrating Axiom 3 above. First, using the tree diagram on the left hand side of

Figure, I obtain $H_1 = -(1/6)\log_2(1/6) - (1/3)\log_2(1/3) - (1/2)\log_2(1/2) = 1.459$. Now, using

the tree diagram on the right-hand side of Figure 1, I obtain $-(1/2)\log_2(1/2)-$
$(1/2)((2/3)\log_2(1/3) - (1/3)\log_2(1/6)) = 1.459$.

One might ask what the original utility of entropy was, in the context of information theory itself. In the same paper, Shannon demonstrates the following theorem, which he calls "The Fundamental Theorem for a Noiseless Channel". (In this context, a "channel" is an arbitrary medium – a coaxial cable, a fiber optic cable, sound waves through air, etc… – through which a source can send a message to a receiver.) The theorem is as follows:

*Let a source have entropy **H** … and a channel have a capacity **C** (bits per second). Then it is possible to encode the output of the source in such a way as to transmit at the average rate **C/H** – ε symbols per second over the channel where ε is arbitrarily small. It is not possible to transmit at an average rate greater than C/H.*

(Ibid., 16)

A few examples will illustrate these concepts.

First, consider the rhythm shown in Figure 3, which represents an excerpt in which the first of every four beats is accented (only accented notes are displayed). This is not an uncommon situation in unsyncopated music. Intuitively, we expect the entropy to be zero, since there is no uncertainty in the IOI's. Mathematically, there is only one non-zero probability, and it must equal one. Therefore, the logarithm in equation 2.1, and thus the entropy, is equal to zero.
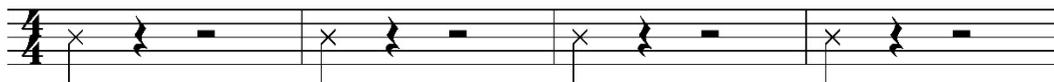


**Figure 3.** A zero entropy rhythm

Next consider the situation illustrated in Figure 4; there is considerably more variation in time delays here. First of all, there are eight different time delays represented here, from one eighth note to an entire measure. Secondly, each of these values occurs with equal probability (1/8). This situation (equal probability for each value) represents the maximum entropy distribution for a given number of possible values a random variable can take on. For $n$ possible values, the maximum entropy is $\log_2 n$, and in this case the entropy is $\log_2 8 = 3$.
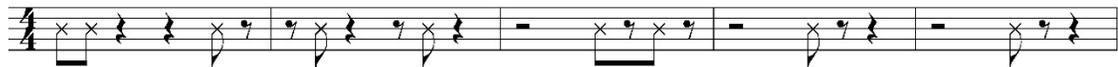


**Figure 4.** A maximum entropy rhythm

Finally, consider the situation illustrated in Figure 5, the Charlie Parker composition "Billie's Bounce"; the distribution of probability of occurrence vs. IOI's is shown in Figure 6. As discussed previously, care must be taken to identify these IOIs; though at first glance it might appear that there is only one IOI – three eighth-notes – between the first seven pairs of accents, some of them start on long eighth-notes and some of them on short eighth-notes. So in Figure 6, note that there are points that fall just to the left of three on the x axis and points that fall just to the right of three on the x axis, reflecting the unevenness of the swung eighth-note pairs. The IOI between the accent on beat four of measure three and the accent on beat one of measure four is two eighth-notes, between the accent on beat one of measure four and the accent on beat three of measure four is four eighth-notes, and so on.

The entropy of this melody is 2.503, and it has seven distinct IOIs. The entropy for seven equiprobable categories of IOI (the maximum-entropy situation for seven IOIs), would be $\log_2 7 = 2.807$. It is not surprising that the entropy for "Billie's Bounce" is less

than the entropy for eight equiprobable IOIs (the example discussed above), since, a) there are fewer IOI categories for "Billie's Bounce," and b) the distribution is not equiprobable (see Figure 5), a requirement for maximum entropy given a fixed number of distinct IOIs.



**Figure 5.** "Billie's Bounce" (head) by Charlie Parker

Some authors define a quantity I will call "normalized entropy," which I define here in the context of IOIs and rhythm. Normalized entropy, for a rhythm having $n$ distinct IOIs, is obtained by dividing entropy by the maximum entropy obtainable for that many IOI's: $\log_2 n$. Thus, for $n$ distinct IOI's, and unnormalized entropy $H$, normalized entropy is given by the following expression:

$$H_N = H / \log_2 n \qquad \qquad \text{Eq. 2.2}$$

Some authors use the term "relative entropy" for this quantity, but this is a misnomer since, according to Cover and Thomas (2006), the term "relative entropy" is reserved for something else (see below).

While normalized entropy gives an idea of how high the entropy of an excerpt is relative to its maximum value, it has a serious flaw. To see this, consider the probability

distribution for the Charleston rhythm (Figure 7). It consists of two IOIs: 3 eighth notes and 5 eighth notes, both of which occur once in each measure. Thus, the entropy is maximized for this many distinct IOIs, and the normalized entropy is:

$$H_N = (- 0.5*\log_2 (0.5) - 0.5* \log_2(0.5))/ \log_2(2) = \qquad \text{Eq. 2.3}$$

$$0.5*\log_2 (2)+0.5*\log_2 (2) = 1$$
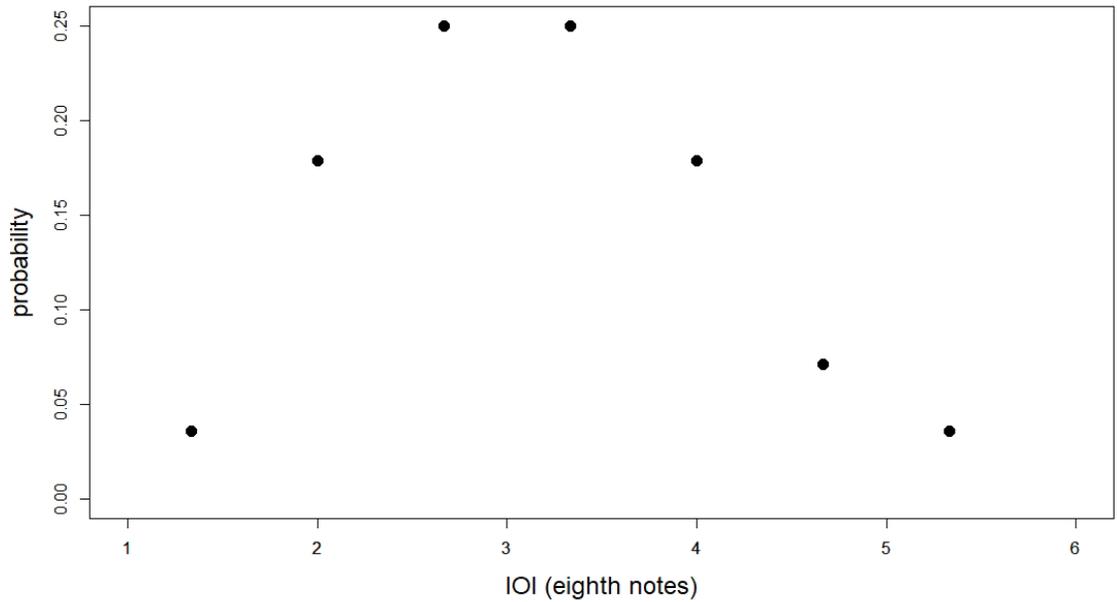


**Figure 6.** Probability distribution for Parker's "Billie's Bounce"



**Figure 7.** The Charleston rhythm

A much more complex rhythm, however, might have substantially lower normalized entropy. For example, the melody of "Billie's Bounce" would have a normalized entropy of just 0.892, since the raw entropy obtained from it is 2.503, and the entropy for seven

equiprobable categories is 2.807. Margulis and Beatty (2008) call this the "rare interval

problem": if one adds just one instance of an IOI not already present in a given

distribution, the denominator of the normalized entropy formula increases substantially,

while the unnormalized entropy increases only slightly and this reduces the normalized

entropy.

Despite the problems inherent in using normalized entropy, many authors use it to

define a related concept, redundancy, using the following formula:

$$R = 1 - H_N$$                                        Eq. 2.4

Redundancy varies between zero and one, with zero corresponding to maximum

normalized entropy, and one to minimum normalized entropy. Intuitively, $R$ represents

the predictability of a probability distribution, though given the limitations of normalized

entropy, it must be used with caution.

In Chapter 5 it will be shown that the issue of normalization can be sidestepped

by evaluating the entropy for each excerpt separately, and making pairwise comparisons

based on entropy between musicians, using a method called "estimated marginal means."


## 2.3 Cross Entropy

A quantity similar to Shannon entropy is "cross-entropy," which can be used to

evaluate the goodness-of-fit of a model to data; Temperley (2007) provides a clear and

intuitive description. Cross-entropy is given by the following expression – similar to the

expression for entropy given above – in which $p$ and $q$ are two probability distributions

defined on the same sample space, corresponding to that of a random variable $X$; $p$

represents data and $q$ represents a model, and the sum is over all values of $X$:

$$CE(p, q) \equiv - \sum_x p(x) \log_2 q(x) \qquad \text{Eq. 2.5}$$

Lower values of cross-entropy indicate a better fit of model to data.


## 2.4 Relative Entropy and Mutual Information

Having discussed entropy, normalization, redundancy, and cross-entropy, I now introduce the concepts of relative entropy and mutual information.

The "relative entropy," also called "Kullback-Leibler distance" or "Kullback-Leibler divergence" (I will use the abbreviation KLD) of two random variables defined on the same sample space with probability distributions $p(x)$ and $q(x)$ is defined as (Cover and Thomas, 2006):

$$D(p\|q) \equiv \sum_x p(x) \log (p(x)/q(x)) \qquad \text{Eq. 2.6}$$

This is not a true "distance," since, for one thing, it is not symmetrical with respect to $p$ and $q$, but it is often useful to think of it as one (ibid.). It will be used here to define mutual information, and to compare certain other distributions to uniform distributions in order to understand differences in entropy.

Several terms must be defined in order to understand mutual information. The elements of the product space $x \otimes y$ are all of the ordered pairs $(a, b)$, where $a$ is drawn from the set comprised of all values of $x$ and $b$ is drawn from the set comprised of all values of $y$. Given a random variable $X$ and a random variable $Y$, their joint probability distribution, defined on the product space $x \otimes y$ and notated $J(x,y)$, is the probability that $X = x$ and $Y = y$ simultaneously. The marginal distributions, $p(x)$ and $q(y)$, give the probabilities that $X = x$ and $Y = y$ individually. The marginal distributions can be calculated from the joint distribution as follows:

$$p(x) = \Sigma_y J(x, y) \qquad\qquad\qquad \text{Eq. 2.7(a)}$$

$$q(y) = \Sigma_x J(x, y) \qquad\qquad\qquad \text{Eq. 2.7(b)}$$

This makes sense because for *p(x)*, we are not interested in values of *y*, but in the

probabilities of *x* regardless of *y*, so for a given *x* we sum over all values of *y* to obtain the

marginal distribution *p(x)*. A similar argument holds for *q(y)*.

Two random variables are considered "independent" if knowledge of the value of

one does not affect our knowledge of the other. Two random variables are independent if

and only if their joint probability distribution is equal to the product of their marginal

probability distributions, in other words, if

$$J(x,y) = p(x)\, q(y) \qquad\qquad\qquad \text{Eq. 2.8}$$

With these concepts in hand, I can now turn to the concept of mutual information.

Given two random variables *X* and *Y* defined on the cross product space *x⊗y* with joint

distribution *J(x,y)* and marginal distributions *p(x)* and *q(y)*, mutual information is defined

as the relative entropy (KLD or "distance") between the joint distribution function and

the distribution that would obtain in the case of complete independence, namely, the

product distribution function of the marginals. In the following formula, *MI(X,Y)* is the

mutual information. Note that it is symmetrical with respect to *X* and *Y*, that is, *MI(X,Y) =*

*MI(Y,X)*:

$$MI(X,Y) = \Sigma_{x\otimes y} J(x, y) log(J(x, y)/\, p(x)\, q(y)) \qquad\qquad \text{Eq. 2.9}$$

$$= \Sigma_x \Sigma_y\, J(x, y) log(J(x, y)/\, p(x)\, q(y))$$

This definition makes sense if mutual information is considered to be the "distance"

between the actual joint distribution and the product distribution. These distributions

would be equal (and hence, the mutual information would evaluate to zero) only in the

case of complete independence; the further from independence the actual joint

distribution is, the more one distribution tells us about the other. Intuitively, mutual

information "is the reduction in the uncertainty of one variable due to the knowledge of

the other" (ibid., 19).


## 2.5 Markov Chains

It is common to take the concept of entropy one step further and look not only at

probabilities of the individual values a random variable takes on in a sequence of

measurements, but also on the probabilities of the individual values *preceded* by one or

more precursors. Probability structures such as these are called "Markov chains." First-

order Markov chains take into account only one precursor, while more generally, Markov

chains of order $n$ take into account $n$ precursors. For a Markov chain of order $n - 1$, the $n$

– 1 measurements preceding a given primary measurement, along with the primary

measurement itself, constitute what is called an "$n$-gram." For example, a first-order

Markov chain would consist of overlapping "bi-grams." For any given Markov chain, a

"probability transition matrix," or "PTM" can be constructed, in which the horizontal

index refers to the preceding measurement(s) in the $n$-gram and the vertical index refers

to the primary measurement itself.


## 2.6 Examples

Before concluding this brief introduction to information theory, I consider five

sample distributions and describe them in terms of their entropy, their Kullback-Leibler

distance (KLD) from a uniform distribution, and their qualitative features. (Recall that

KLD is not symmetrical with respect to the two distributions whose distance it measures; thus, in what follows, KLD will be calculated three ways: once for each disposition of the sample and uniform distributions, and averaged between the two). All five have ten different IOI's and 31 notes (thus 30 IOIs total).

The first distribution is the uniform distribution with 10 IOI categories (not shown), which has the maximum entropy for 10 categories of $\log_2 10 = 3.322$. Obviously, the KLD of this distribution compared to the uniform distribution is zero, since the distance from a given distribution to itself is zero.

The second distribution (Figure 8) has a single peak – corresponding to twelve instances of IOI "one" – and two instances each of IOIs 2–10. This has a mean KLD from the uniform distribution for 10 IOIs of 0.269, and the entropy is 2.873.
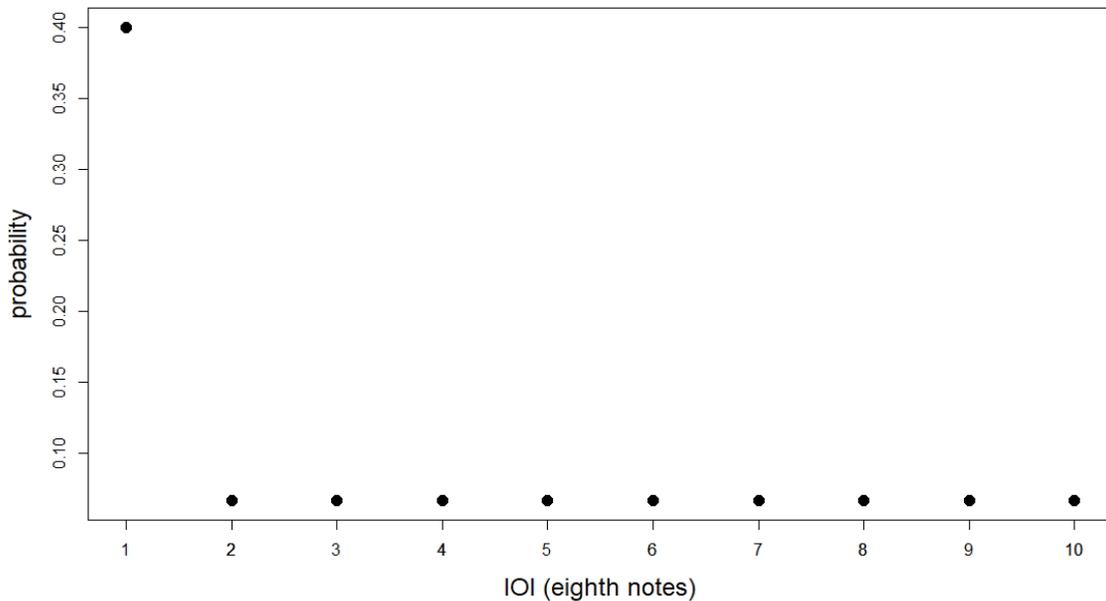


**Figure 8.** The second example distribution described in the text

The third distribution (Figure 9) has three peaks of roughly equal magnitude (corresponding to nine instances of IOI "one," and seven instances each of IOI "two" and IOI "three"), and one instance each of IOIs 4–10. The KLD has mean value 0.479 and the entropy is 2.646.

The fourth example (Figure 10) has two peaks of magnitude equal to approximately three halves the magnitude of the three peaks in the third distribution; in other words, the amount of probability accounted for in the two peaks in this example is about the same as the amount of probability accounted for in the three peaks in the third distribution. The mean KLD is 0.639 and the entropy is 2.37.
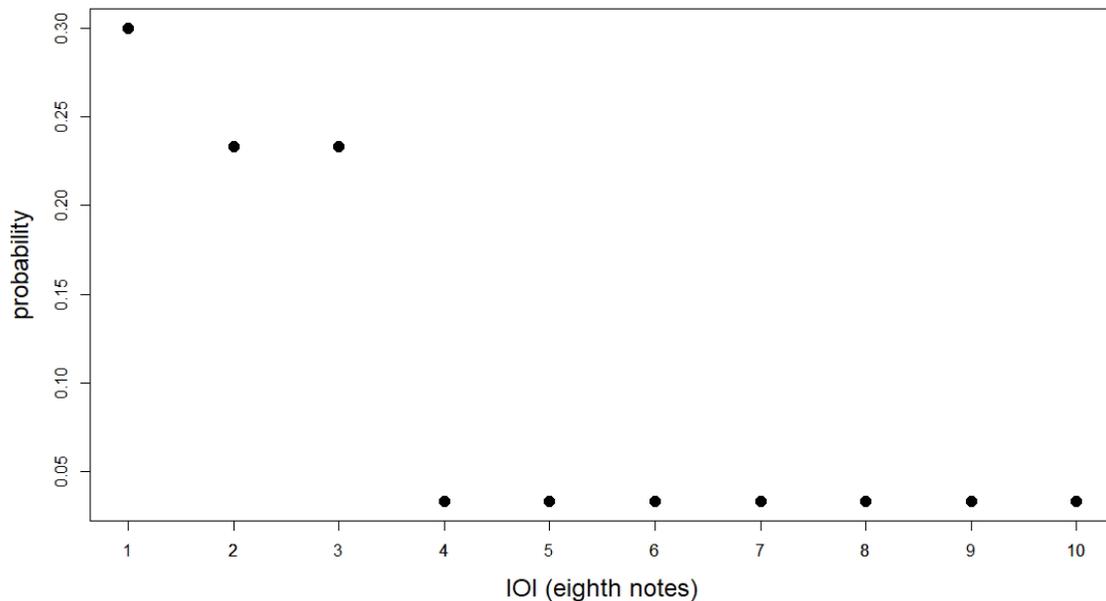


**Figure 9.** The third example distribution described in the text
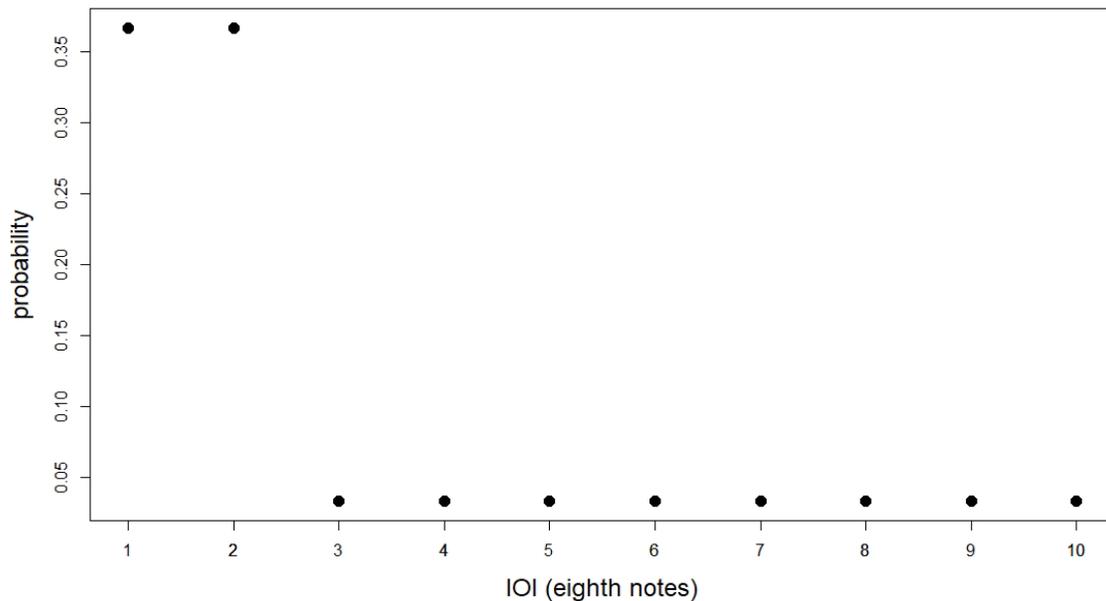
**Figure 10.** The fourth example distribution described in the text

The fifth and final distribution (Figure 11) has a single peak accounting for about the same amount of probability as the combination of the two peaks in the fourth example. It has a mean KLD of 0.913 and an entropy of 1.832.

These sample distributions have been arranged in order from least to greatest KLD, which, here, is the same as arranging them from greatest to least entropy. This is usually, but not always, the case. It is not surprising that entropy is usually inversely correlated with KLD from the uniform distribution, since for a given number of distinct IOIs, the uniform distribution always has the greatest entropy.

This also tells us that, qualitatively, there are two features in particular that affect entropy: number of peaks and amount of probability accounted for in those peaks. I have shown that a distribution with a single small peak (not accounting for much probability) has a lower KLD (and higher entropy) than a distribution with three small peaks.

However, deleting peaks one at a time from the distribution with three small peaks, while maintaining the amount of probability accounted for in those peaks, *increases* KLD (and *decreases* entropy). So the distribution with two moderately sized peaks has higher KLD and lower entropy than the one with three small peaks, and the distribution with a single large peak has a higher KLD and lower entropy still.
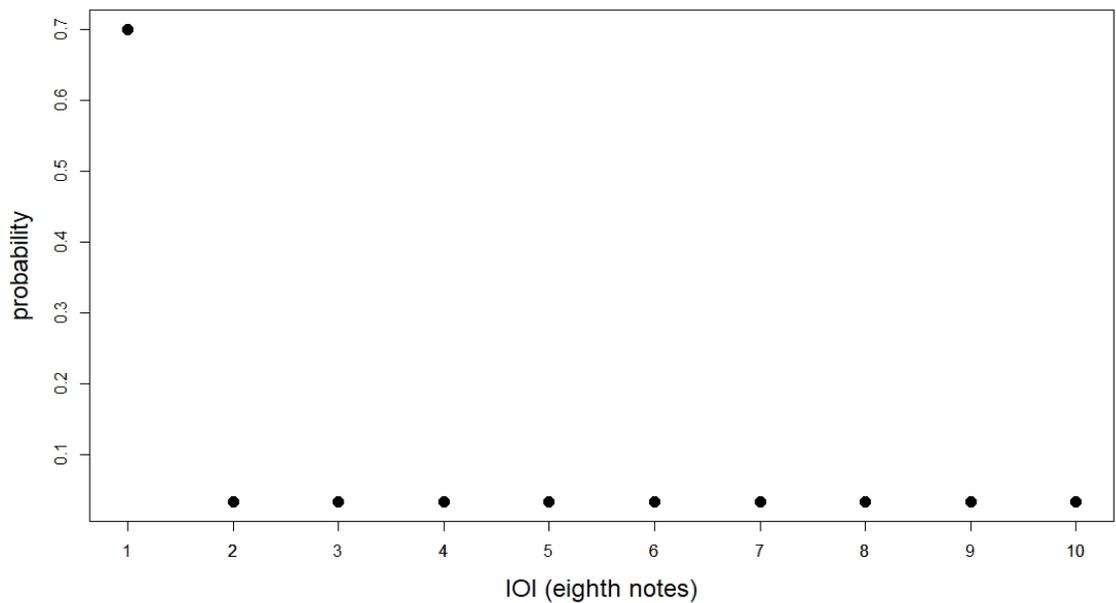


**Figure 11.** The fifth example distribution described in the text

These trends make sense when we consider an alternative definition of entropy, as the average "surprisal" of a distribution, where the surprisal of an event *E* with probability *P(E)* is log(1/*P(E)*), which is the same as – log(*P(E)*). (The surprisal of an event is the same as its "information content.") Note that taking the average of the surprisal does amount to calculating the entropy, since the average value of any function *f(x)* is:

$$\sum_x p\,(x)f(x) \qquad \text{Eq. 2.10}$$

23

For a distribution with one sharply defined peak, the average surprisal is low, since the same event occurs with very high frequency and thus has a low surprisal value – and contributes most to the sum used in computing the average surprisal value. As peaks are added, the average surprisal value increases until, at last, we are left with a uniform distribution, which has the highest average surprisal and thus, entropy, possible with a given number of IOI categories.

The second distribution above bucks this trend: it has only a single peak, but a high entropy compared with the three distributions that follow. This peak, however, accounts for less probability than the peaks in the other distribution. That is why it does not follow the trend.

We will encounter all of these archetypical examples in the chapter on close readings.

## 2.7 Terminology

The term "entropy" is used extensively throughout this document. At times it is used in the abstract sense of Equation 2.1, but more often than not it refers to the more specific case of "entropy calculated from the distribution of IOIs between accented notes in jazz solos." Context should make it clear which definition applies in a particular case.

# CHAPTER 3

# A REVIEW OF THE LITERATURE

## 3.1 Introduction

The techniques used in this research fall under the heading of information theory, which in turn falls under the heading of probability theory. Information theory is built on related concepts including entropy, normalized entropy, redundancy, cross-entropy, relative entropy, mutual information, Markov chains, and surprisal. These concepts were introduced in Chapter 2. Elements of probability theory not directly related to information theory are also frequently used in the study of music.

The intersections of information theory and probability theory with music will be discussed below. Subsequent to that discussion, studies of rhythmic complexity (including metrics based on the abstract concept of entropy), an overview of jazz and Markov chains, and a brief history of the evolution of jazz rhythm will be presented.

## 3.2 Information Theory, Probability Theory, and Music

Within a decade after the publication of Shannon's 1948 paper, authors had begun applying information theory to music, with results both practical and theoretical.

A popular science article by Pinkerton (1956) uses redundancy (as defined in Chapter 2) to analyze folk melodies, and introduces the concepts of the first-order Markov chain and the first-order probability transition matrix (PTM – see Chapter 2) as applied to music. He suggested that harmony and rhythm can be similarly analyzed using entropy in future work, and demonstrated how a first order PTM can be used to randomly compose simple songs.

To compose music using a first order PTM, one chooses a pitch to begin with. One then looks up the probability of each pitch following the starting pitch, including perhaps a certain probability for there to be a rest. Based on these probabilities, one selects the next note or rest. One then reiterates this procedure, using each newly selected note (or rest) as a starting event, until an end is chosen.

Meyer (1957) posits that concepts central to the science of musical meaning, as defined in his *Emotion and Meaning in Music* (1956), have "striking parallels – indeed equivalents – in information theory" (Meyer 1957, 412). The fundamental reason for this is that musical meaning, according to Meyer, is created by the inhibition of tendencies for given musical events to resolve as the listener expects them to; this can be modelled using information theory. The simplest mathematical structure with which to model antecedent-consequent relationships is the first-order Markov chain. If the probability of a consequent following a given antecedent is high, the information content, and thus the meaning, of that consequent is low; if the probability of a consequent following a given antecedent is low, the information content, and thus the meaning, is high. Note that the concept of a probabilistic model is central both to information theory and to the science of musical meaning: a tendency can only be inhibited if it has a high probability of occurring. Overall, he showed that "…meaning and information arise out of the same processes." (Meyer 1957, 423)

The first to analyze melodies from the common practice period using entropy and related tools, and thus, the first to realistically analyze actual music using information theory, was Joseph Youngblood. In a 1958 paper, he states:

*[o]ne of the things which a piece of music communicates is information about the style. A certain amount of redundancy is necessary for the transmission of this information. Rather than seeking to eliminate this redundancy, we are interested in finding and measuring it.*

<div align="right">(Youngblood 1958, 29)</div>

Working within a mod12 pitch-class space, he transposes each piece so that its structural (global) tonic is zero. He then computes entropy and normalized entropy by tabulating the across-corpus frequencies with which each pitch-class occurs. He uses normalized entropy to calculate the redundancies of pieces by Mendelssohn, Schumann, and Schubert, and finds the greatest redundancy in Mendelssohn. However, the differences between the three composers are small (<17%), and in the absence of further analysis, cannot be said to be statistically significant or not.

He also examines excerpts of Gregorian chant, working within a mod7 space, and finds *lower* redundancy than in the case of the Romantic composers. However, when he considers these excerpts within the context of *twelve* possible pitch classes, he finds *greater* redundancy than in the Romantic composers, except for Mendelssohn.

He calculates probability transition matrices for both the Romantic composers and the Gregorian chant. For each row of these matrices (corresponding to each antecedent note) he calculates entropy, normalized entropy, and redundancy, and subsequently averages those quantities over all possible antecedent notes (namely, the 12 pitch classes). Finally, he averages over all elements of the PTM.

Rather than simply applying the concept of entropy directly to music, Coons and Kraehenbuehl (1958) propose an intriguing but unwieldy alternative metric. Their metric

is intriguing because it reflects the moment-to-moment experience of a subject listening to a piece of music: it is evaluated by classifying each new event as similar or dissimilar to what has come before. It is unwieldy because the formulae used to compute their metric is complex, as opposed to entropy, which can be calculated from a simple closed-form mathematical expression.

In a 1962 article in *Behavioral Science*, Joel Cohen explores numerous aspects of how information theory can be applied to music. First of all, he points out that many authors take mathematical liberties with information theory in applying it to music. He posits that five conditions must be met – and seldom are – in order for information theory to be applied to a given musical sample accurately (Cohen 1962, 155): stochasticity (the sample can be well-described by a random probability distribution); ergodicity ("a sufficiently large sample from an infinite sequence has the same statistical structure as the infinite sequence"; stationarity (the probability distributions in question do not depend on time); Markov consistency (the same order Markov sequence prevails throughout); and infinite memory capacity (which, of course, never holds, but can be approximated).

Cohen posits that composition may be regarded as "selecting acceptable sequences from a random source" (ibid. 147). He discusses computer composition, the music of Cage, and the music of the total serialists. He states that the music of the total serialists should be 100% redundant but in practice is not (ibid. 147). He cautions against the haphazard application of entropy calculations to problems in music theory. Finally, he raises four question that may indicate ways forward for the study of information theory and music (ibid. 159):

*What can an information theory of music say about the emotional or affective response to music? What can it say about the psychological sign system of music? How is it related to the general problem-solving of human activity? What, in particular, are its effects on the hoary domain of aesthetics?*

*Cohen 1962*

The present work deals with information theory and *rhythm*; to my awareness, the first studies to deal explicitly with information theory and rhythm appeared in the late 'fifties and the mid 'sixties. The earliest was an unpublished 1959 Indiana University dissertation by J.G. Brawley, Jr. which I did not obtain.

A second early precedent for the application of information theory to rhythm is the work of Hiller and Fuller (1967) on Webern's *Symphony* Op. 21. In it, the two authors analyze inter-onset intervals (IOI's), in addition to pitch, using a variety of entropy-based metrics, including: average information content (entropy) per "measurement" of a random variable representing IOI or pitch; average information content per two-element combination of values of that random variable; second-order calculations, evaluating the averages of conditional probabilities, i.e., probability of event i given event j; all higher order extrapolations of these fundamental metrics; and redundancy. They examine redundancy as it depends on section (exposition, development, and recapitulation), and find that the exposition has the simplest IOI structure, while the recapitulation has the most complex IOI distribution.

Following Hiller and Fuller's article, there was a hiatus in information theory and music articles until the early 1980's. Beginning with an article by Knopoff and

Hutchinson (1981) and continuing with articles by Knopoff and Hutchinson (1983) and

Snyder (1990), authors began to refine previously used techniques.

Knopoff and Hutchinson (1981) investigate the question of how the analysis of

entropy in continuous data differs from the analysis of entropy in discrete data. They find

a quantity they call absolute entropy that depends only on the "bin size" used to discretize

continuous data, a quantity that cannot be calculated but that can be ignored if

comparisons between calculations using the same bin size are made. They apply their

method to the analysis of acoustic dissonance, and find that the histogram of the

logarithm of acoustic dissonance very closely approximates a normal distribution. They

introduce the concept of the "bootstrap method" – though they do not use this term – for

calculating confidence intervals, and apply it to the calculation of the entropy of acoustic

dissonance in Bach Chorales[2]. Finally, they demonstrate the use of durational weighting

in the calculation of the entropy of dissonance.

The same authors, in a 1983 article, take up the concept of bootstrap confidence

intervals again, and use them to show that increasing the sample length by a factor of $N$

decreases the percentage spread of entropy by a factor of $\sqrt{N}$. They demonstrate that

some of Youngblood's conclusions were false because the confidence intervals implied

by his data overlapped. They present some data of their own, using confidence intervals,

and demonstrate, for example, that the entropies for Mozart and Johann Adolph Hasse are

not significantly different at the 95% confidence level, but that the entropies for Schubert

and Richard Strauss are different from Mozart/Hasse and from each other. They arrive at

---

[2] Bootstrapping – in this case, nonparametric bootstrapping – is a method in which the sample distribution
is used as a stand-in for the population distribution, and a large number of "bootstrap samples" are drawn
from it and used to calculate confidence intervals or $p$ values.

an approximate empirical formula for confidence intervals in the case of normalized entropy for pitch distributions: $H_N$ +/-3.65/ $\sqrt{n}$, where $n$ is the number of points in a random sample. This formula appears to be valid "to a very high accuracy … independent of historical period".

As a further exploration and refinement of previous techniques, Snyder (1990) critiques several assumptions made by previous analysts – including Youngblood, Cohen, and Knopoff and Hutchinson – regarding "relative pitch class," "key signature dependen[ce]," and "modally biased tabulation". Under relative pitch class assumptions, a 12 pc universe is used, but all pitch classes are related to the local tonic. Under key signature dependence assumptions, the only modulations taken into account are those which are notated in the score. Using modally biased tabulation, all major mode melodies are transposed to C major, while all minor mode melodies are transposed to A minor (thus modally biased tabulations could be called "la-minor" tabulations). Snyder, by contrast, advocates using scale degree assumptions, structural tonic assumptions, and "do-minor" tabulations. Under scale degree assumptions, the alphabet for pitch has more than 12 elements, obtained from the diatonic scale degrees and their inflections. Under structural tonic assumptions, following Schenker, all pitches are related only to the structural tonic. Using do-minor procedures, major mode melodies are transposed to C major and all minor mode melodies are transposed to C minor. He demonstrates that the approach he advocates yields substantially different results from the approach used by his predecessors, but that the differences are difficult to predict. He does state, however, that they are "more discriminating" and "more plausible" than results obtained using the earlier assumptions. Finally, he advocates five routes for further exploration: 1)

durational weighting of pitches; 2) deciding whether or not to count "reiterations" of pitches; 3) treatment of verses in strophic songs; 4) analyzing other structures such as chords, and; 5) higher-order Markov chains (entropy can be considered to be the result of calculations based on "zeroth order" Markov chains).

Note that Youngblood, Knopoff and Hutchinson, and Snyder grouped samples from different pieces by the same composer together before calculating entropy and comparing entropy between composers. The current study takes a different approach that will be discussed in Chapter 5.

Again there appears to be a hiatus between 1990 and 2002 in articles pertaining to information theory and music; following Snyder's 1990 paper, Samer A. Abdallah wrote a dissertation – not summarized here – in 2002 titled "Towards music perception by redundancy reduction and unsupervised learning in probabilistic models." The following two decades saw a number of publications on the subjects of information theory and/or probability theory applied to music.

Pearce and Wiggins (2004) evaluate several techniques for improving the performance of *n*-gram models as applied to monophonic music. They compare these techniques by applying them to eight monophonic data sets, and use cross-entropy with 10-fold cross-validation to show that "significant and consistent improvements in performance are afforded by several of the evaluated techniques."

Going beyond the limits of information theory, Sadakata, Desain, and Honing (2006) explain the discrepancy between experimental results based on perception of rhythms and production of rhythms – that is, between the ease with which a subject identifies a given rhythm or reproduces that rhythm – using a Bayesian approach.

Bayesian theory is based upon a theorem called Bayes' Theorem, which in music takes the following form:

$$P(M|S) = (P(S|M)P(M))/P(S) \hspace{4cm} \text{Eq. 3.1}$$

where $P(M|S)$ is the "probability of a model given a musical surface," $P(S|M)$ is the "probability of a musical surface given a model," and $P(S)$ and $P(M)$ are the "probability of the surface" or "prior probability" and the "probability of the model" or "marginal probability" respectively. In the musical context, "prior probabilities" are the probabilities of particular models appearing in the literature. Prior probabilities are usually assumed to be uniform, but Sadakata et al. explore other ways of calculating priors as well: uniform, theoretically derived from a Farey tree, derived from frequencies found in three corpora, and optimized. They find that using non-uniform priors greatly improves the match between perception and production data.

While not introducing techniques that were actually *new* to music theory, Huron (2006), encouraged the adoption of probabilistic methods (including information theory techniques) in studies that followed. Furthermore, Huron's book focused on perception experiments, which form an important component of the present study.

Huron's 2006 book, *Sweet Anticipation*, discusses many aspects of probability theory as it applies to music. This is not surprising, given that the phenomenon of *anticipation* in the title relies on probability and statistics for its meaning: we can only *expect* a musical event in a given context if we have heard it many times before, and thus experience a high probability that it will occur again. (This is related to the discussion of meaning in music above and its inverse correlation with probability).

He uses statistics to model several aspects of music perception, including tonality and rhythm. For example, he resolves a conundrum in Krumhansl's and Kessler's (1982) work on key profiles, in which subjects listen to a key-defining context and then judge how well a subsequent probe-tone fits with the key-defining context. The resulting distribution approximated but did not exactly match the frequency of scale degrees in the literature. Huron and his student, Bret Aarden, were able to demonstrate that the discrepancy is due to the difference between scale tone *frequency* distributions and scale tone *closural* distributions.

Regarding rhythm, Huron uses information theory to compare the information content in a passage from *The Rite of Spring* to the information content in a random rhythm to demonstrate that Stravinsky's rhythms have far more information. For zeroth-order calculations (entropy), the result for a random rhythm was 1.73 (bits) and for Stravinsky was 2.31. For first-order (Markov chain) calculations, the result for a random rhythm was 1.32 and for Stravinsky was 6.64. Huron also reports the research of Patel and Daniele (2003), discussed below.

Overall, Huron posits that "the tools of empirical research have advanced sufficiently" to enable music theorists to look beyond form and explain function.

A book by Temperley (2007) gives an overarching view of Bayesian theory applied to music. He presents the basic concepts of Bayesian theory clearly and uses them to derive a meter-finding algorithm for monophonic music, a key-finding algorithm for monophonic music, a polyphonic key-finding algorithm, and "Bayesian models of other aspects of music". He applies the monophonic algorithms to the problem of error detection, and the polyphonic model to problems related to characteristics of tonality in

music of the common practice period. Finally, Temperley uses cross-entropy to make comparisons between some simple models of pitch and rhythm, to make inter-stylistic comparisons, and to perform some simple tests of Schenkerian theory.

A further evaluation and refinement of previous techniques is provided by Margulis and Beatty (2008), who identify three obstacles to applying information theory to music: obscurities in the framing of musical questions; practical obstacles to tabulating musical entities; and uncertainties regarding the type of musical entity that should serve as the unit of analysis. They use scores in the **kern format using the Humdrum toolkit in order to address the latter two of these difficulties. They use entropy, conditional entropy, and normalized entropy (as discussed in Chapter 2) to analyze several corpora: Bach chorales, Bach preludes and fugues, barbershop quartets, Corelli trio sonatas, Handel trio sonatas, Telemann violin sonatas, Haydn string quartets, and Mozart string quartets. They refer to the work of Knopoff and Hutchinson – in particular, their claim that 7,900 characters are necessary to distinguish results from statistical noise, and select their corpora to satisfy this condition. Using these metrics and corpora, they examine eight parameters related to pitch, interval, contour, and rhythm, and draw numerous conclusions, including most importantly: 1) "the central hypothesis that most styles would "specialize" in a certain parameter, concentrating much of their entropy in a particular dimension"; and 2) that there is a finding of "a gradual accrual of entropy across time." They point out the "rare interval" problem, namely the fact that an interval (or other musical parameter) that occurs only rarely increases the unnormalized entropy only slightly but drastically decreases the denominator in the formula for normalized entropy, thus decreasing the normalized entropy (see discussion of this phenomenon in

Chapter 2). Finally, they suggest using a model of perception in conjunction with cross entropy rather than entropy alone to measure musical structure.

Revisiting the topics presented in his book (2007), Temperley (2010) uses Bayesian reasoning, the technique of cross-entropy, and the concept of minimizing model complexity to evaluate the performance of six probabilistic models of rhythm applied to two corpora of European music – the Essen folksong corpus and a corpus of first violin parts for all of Haydn's and Mozart's string quartets – and identifies a model that places second in terms of cross entropy but first when model complexity is taken into account. (The technique of weighing model performance against model complexity will be returned to later on). Again using Bayesian reasoning, he tests these models as models of perception rather than production, focusing in particular on their ability to find meter and to quantify syncopation, and identifies the same two models that performed best as models of production as the best models of perception, though more work is needed.

Finally, moving beyond Bayesian theory, Temperley explores the issue of "communicative pressure," that is, the desire on the part of the composer to "communicate" musical information to the listener. As an example, he cites the "desire" on the part of the composer of counterpoint to project the sound of independent voices. This creates the pressure to avoid parallel perfect intervals and to avoid small intervals in low registers, as these would mask the independence of the voices. He also discusses the concept of "trading relationships," which describes how when one musical parameter changes, another changes to compensate for the change in the first one. As examples, he cites rubato/syncopation, syncopation/swing (which, he argues, are compensatory, as

swing is an indicator of strong eighth vs. weak eighth), and the use of upper extensions in jazz/the freer use of chordal inversions in common practice music.

Some authors use the information content of individual events rather than in the aggregate (i.e. rather than entropy); Duane (2012) is one such author. Duane (2012) discusses anthropomorphized virtual "agents" in music, which can be either *leading* or *subordinate*. He identifies a type of sonata exposition in which a texture characterized by a single, leading, agent in the first theme gives way to a texture involving two leading agents, and calls this expositional paradigm an "emergent second agent" paradigm. He shows that this type of exposition can be differentiated from expositions *not* characterized by emergent second agents on the basis of a complicated calculation involving the information content (surprisal) of individual events (see Chapter 2), and therefore that listeners may be influenced by information content in the course of perceiving virtual agents.

Three-chord trigrams serve as the basis for several experiments conducted by White (2014). First, he tallies the frequency of occurrence of thirty trigrams centralized to a common pitch level and constructs a vector of trigram frequencies for each of nineteen composers. He performs a cluster analysis based on the cosine of the angle between each pair of vectors, and finds that the vectors are indicative of style, interpreted chronologically, geographically, and in terms of composer influence. Still using these vectors, he calculates cross entropies between each pair, and finds that chronological distance is reflected in cross entropy values. He plots his results using the Handel corpus as his reference point, and finds that a straight line predicts the result with a few exceptions. The final experiment consists of using corpus-sensitive key-finding

algorithms to find the key in the first few measures of the *Eroica* Symphony; he finds that algorithms trained on different corpora come up with different key identifications.

Jacoby et al. (2015) study "theories" that group surface "tokens" into categories, and evaluate these theories according to accuracy and complexity. "Tokens" are any objects that can be used to describe chords at the musical surface, such as Roman numerals or chord quality descriptions. They use entropy, among other metrics, to measure complexity; they use mutual information (Cover and Thomas 2006) among other metrics, to measure accuracy. They use the method of Lagrange multipliers to maximize accuracy while minimizing complexity. They construct an "evaluation plane" by plotting accuracy against complexity, and plot an "optimal curve" for models on that plane. Surprisingly, they find that fundamental bass theory is slightly more accurate than root-functional theory. However, the difference in accuracy is small (.05 bits), while the difference in complexity is large (0.47 bits), so the root-functional theory is preferred over the fundamental bass theory. This method is also applied to Palestrina and rock.

Another author who works with the information content of individual events is Temperley (2019), who explores the "theory of Uniform Information Density". This theory states that "communication is optimal when information is presented at a moderate and uniform rate". He proposes three characteristics of music that satisfy this theory, and tests them against Renaissance counterpoint, expressive performers, and common practice themes.

The current work lies at the intersection of information theory and rhythmic complexity. Not all information theory is used for the purpose of gauging rhythmic complexity, and not all gauges of rhythmic complexity are based on information theory.

In the section that follows, I will discuss a number of rhythmic complexity measures, some – but not all – of which rely on information theory.

## 3.3 Rhythmic Complexity

The concepts of rhythmic complexity and syncopation are distinct but not without commonalities. Roughly speaking, syncopation is a kind of rhythmic complexity, one that relies on the metrical placement of notes for its definition; rhythmic complexity may or may not depend on meter. In this section, measures of syncopation and rhythmic complexity more generally will be discussed, and attention will be paid to how these metrics model rhythmic complexity and how they are behaviorally tested.

Longuet-Higgins and Lee (1984) present a simple but elegant metric for describing syncopation. Since this will be applied in the experimental portion of this study (Chapter 4), I will explain it here in some detail.

The LHL metric, when applied to eighth-note rhythms in 4 4 time, works by assigning each part of the measure a number: 0 for the downbeat, –1 to beat three, –2 to the off–beat quarter notes (two and four), and –3 to the four off-beat eighth notes (the "ands" of one, two, three, and four). A syncopation is said to occur when a note sounds before a tied note or rest having a greater (more positive) value than the sounding note. The "weight" of the syncopation is equal to the greater value minus the lesser value, which will necessarily be a positive number. Note that a single sounded note can correspond to multiple syncopations, according to the number of rests and/or tied notes following the sounded note.

Several examples are shown in Figure 12: (a) A note that sounds on the "and of two" (i.e., its note attack begins an eighth–note after the second beat) and is tied to a quarter note on beat three will count as a syncopation, since the value assigned to the tied note (–1) is greater than the value assigned to the sounding note (–3). The weight of the syncopation will be –1 – (–3) = 2. If there is a rest on beat four, there will be an additional syncopation of weight 1, obtained from the difference between the value assigned to beat four (–2) and the value assigned to the sounding note (–3). (b) A quarter note that sounds on beat four and is tied over to the beginning of the next measure will count as a syncopation, since the value assigned to the tied note (0) is greater than the value assigned to the sounding note (–2). The weight of the syncopation will be 0 – (–2) = 2. (c) A note that sounds on the and of four and is tied to the beginning of the next measure will count as a syncopation with weight 0 – (–3) = 3. If there is a rest on beat two, there will be another syncopation, this one of weight 1, obtained from the difference between the value assigned to beat two (–2) and the value assigned to the sounding note on the and of four (–3).
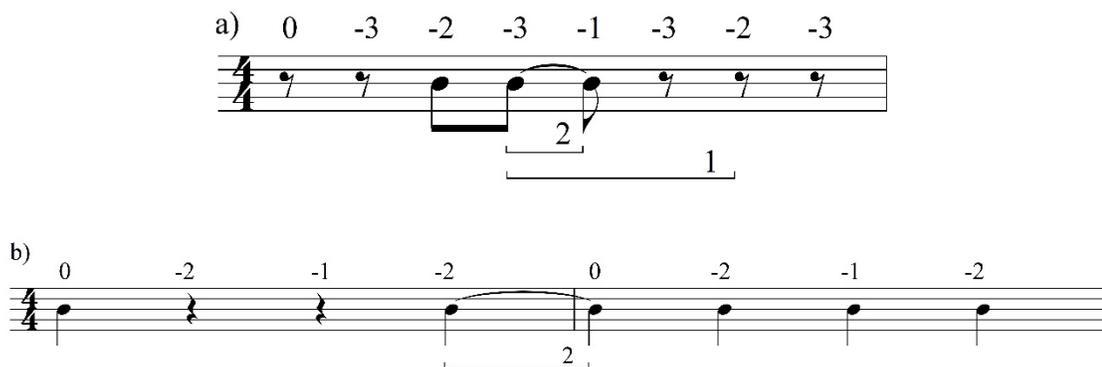


**Figure 12a, b.** Two out of three examples of the LHL metric: a) a total of three (two syncopations initiated by the same note); b) a total of two (one syncopation only)
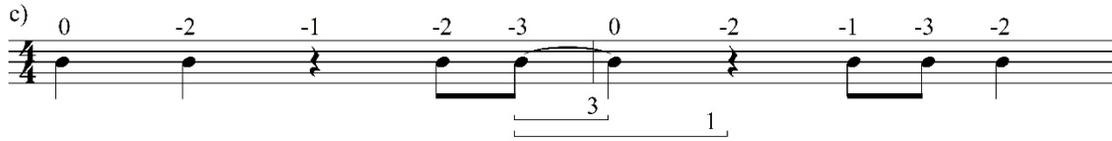
**Figure 12c.** A third example of the LHL metric: a total of four (two syncopations initiated by the same note)

A set of experiments by Povel and Essens (1985) studied the effects of internally induced clocks (akin to meter) on the ease of perceiving test rhythms. Their first experiment shows that listeners found it easier to reproduce test rhythms that strongly induced an internal clock. The second experiment shows that adding a metronomic pulse aligned with the induced clock made it easier still to reproduce the test rhythms. Finally, adding a metronomic pulse that does *not* align with the induced clock makes it *harder* to reproduce the test rhythms.

Building on this study, Shmulevich and Povel (2000) discuss three quantitative measures of rhythm complexity: the Tanguine "*T* measure," Lempel-Ziv complexity, and the Povel-Shmulevich measure, the latter based upon the work of Povel and Essens. They test these models by correlating them with subjective rankings in an experiment in which participants were asked to rank the complexity of thirty five rhythms on a five-point scale. One of their measures, the Povel-Shmulevich measure, was found to correspond much more closely to the subjective rankings than the other two.

Smith and Honing (2006) use the data from the experiment of Shmulevich and Povel (2000) to evaluate the syncopation metric of Longuet-Higgins and Lee (1984). In fact, they extend this metric by using three different metrical weighting schemes: the scheme originally proposed by Longuet-Higgins and Lee, and two proposed by Palmer and Krumhansl (1990), one which applies to musicians and one which applies to non-

musicians. They find that the original weighting schemed proposed by Longuet-Higgins and Lee, and the scheme proposed by Palmer and Krumhansl for musicians correlate highly with the experimental data ($r = 0.75$ and $r = 0.76$ respectively), while the scheme proposed by Palmer and Krumhansl for non-musicians correlates with the experimental data less strongly ($r = 0.64$).

A new experiment designed to test the metric of Longuet-Higgins and Lee – extrapolated to include sixteenth-note subdivisions – was carried out by Fitch and Rosenfeld (2007). Their experiment asked participants to tap along with rhythms ordered by complexity as measured by the metric (more specifically, ordered by amount of syncopation), and to reproduce the rhythms both immediately and after a twenty-four hour delay. Not surprisingly, participants had an easier time tapping along with and reproducing less syncopated rhythms.

Tan, Lustig, and Temperley (2019) improve upon Longuet-Higgins and Lee's "positional" method of measuring syncopation – in the case of vocal music – by accounting for stressed and unstressed syllables implied by the lyrics; they also point out that in rock, most syncopations are anticipations. They introduce several quantitative measures of anticipatory syncopations and use them to compare rock to 19th c. English vocal music and to study the dependence of syncopation in rock on chronology and tempo.

Correspondences between speech and music are explored by Patel and Daniele (2003) using a metric called the "normalized pairwise variability index," or nPVI, that reflects the variability between successive pairs of durations in speech or in music. They demonstrate that instrumental music from England shows more variation than music from

France, in keeping with parallel results in English and French speech. Temperley (2017), however, expanded their corpus to include German and Italian music and language in addition to English and French. Unexpectedly, he did not find a positive correlation between nPVI for music and speech; in fact, he found the opposite: the English and German languages have a higher nPVI for speech than the Italian and French languages, but a lower nPVI for music. Two subsequent papers apply the nPVI metric to music only. Toussaint (2012, 1001), finds that, while the nPVI "suffers from several shortcomings," it "[n]evertheless correlates mildly, but significantly, with human performance complexity." Toussaint (2013, 2–3), introduces a modified nPVI metric and find that it "correlates highly with two measures of human performance complexity for music which is syncopated."

Huron and Ommen (2006) use a different group of quantitative measures of syncopation to test and prove two hypotheses: (1) The density of syncopated moments in syncopated American popular music increased throughout the early years of the twentieth century, and (2) The variety of syncopated patterns in this music increased throughout the early years of the twentieth century.

A survey of eight measures of rhythmic complexity (which the authors call "syncopation", while acknowledging that the term is used loosely) is provided by Gomez et al. (2007). These measures are compared to each other and to the experimental data of Shmulevich and Povel (2000). The authors calculate a rank correlation matrix using Spearman rank correlation and find that, with the exception of two tests (Pressing's measure of cognitive complexity and the rhythmic oddity property), all measures of complexity have statistically significant correlations with one another and with the

experimental data of Shmulevich and Povel. Cluster analysis reveals interesting patterns as well.

Thul and Toussaint (2008) review no fewer than 32 measures of rhythmic complexity, most importantly for present purposes, an adaptation of Shannon entropy. They use frequencies of inter-onset intervals to evaluate entropy, and find that it does not correlate highly with human performance measures, nor with syncopation measures. Note that to measure syncopation they extrapolate the method of Longuet-Higgins and Lee to include sixteenth-note subdivisions. These results will be addressed in Chapter 4.

A further pair of studies involving rhythmic complexity metrics based on information theory is provided De Fleurian et al. (2014, 2017). In particular, these authors evaluate five information-based approaches to complexity: Shannon entropy, entropy rate, excess entropy, transient information, and Kolmogorov complexity. They do so by asking subjects to listen to sequences of notes and rests and, at the end of each sequence, to decide whether a note or a rest should follow, and to rank the difficulty of making that decision. They calculate each of their metrics by representing their test rhythms with strings of one's and zero's, one's for note onsets, zero's for everything else[3]. They find that only two of the measures they tried predicted their experimental results well: entropy rate and Kolmogorov complexity.

The theory of syncopation in rock is taken up by Temperley (1999), who adds to the theory in two ways. For songs with lyrics, he adds a "metrical preference rule" to the

---

[3] Note that this method of representing a rhythm must be used with caution, because changing the granularity to accommodate for finer levels of subdivision, changes calculated measures such as entropy. As long as excerpts with the same level of subdivision are compared, there is no difficulty.

rules of Lehrdal and Jackendof (1983) that states "Prefer a metrical structure in which strong beats coincide with stressed syllables of the text". And he posits the existence of an unsynopated "deep structure" for any given surface that exhibits syncopation, from which the surface can be derived by shifting a syllable from a strong beat to a weak one.

A further study of syncopation by the same author (Temperley, 2019) focuses on "second position" and "fourth position" syncopations, that is, syncopations that occur on the second eighth of a four eighth-note pattern or the second sixteenth of a four sixteenth-note pattern. Temperley cross-tabulates their frequency in English, Scottish, Italian, French, German, Euro-American and African-American songbooks from the nineteenth century. Overall, he finds that second-position syncopations are more prevalent in English and Scottish songs than in continental European songs, and that second-position syncopations are about equally prevalent in British and Euro- and African-American songs, while fourth-position syncopations are more prevalent in African-American song than in British song.

## 3.4 Measures of Rhythmic Complexity: A Summary

Since I am interested in exploring the effects of entropy on perceived rhythmic complexity, I will summarize fifteen ways of quantifying rhythmic complexity discussed in the literature reviewed so far. There are certainly others; as stated previously, Thul and Toussaint (2008) identify and test 32 of them. Fifteen will be considered enough to provide a useful benchmark here. They are:

1) The Tanguine measure, which "uses the idea that a rhythmic pattern can be described in terms of (elaborations of ) more simple patterns, simultaneously at different levels" (Shmulevich and Povel 2000, 4)

2) Lempel-Ziv complexity, "which is related to the number of steps in a self-delimiting production process by which … a sequence is presumed to be generated" (ibid.)

3) The Povel-Shmulevich measure, which "takes into account the ease of coding a temporal pattern and the (combined) complexity of the segments resulting from this coding" (ibid.)

4) The Longuet-Higgins and Lee metric, with the original weighting scheme (see above) (Longuet-Higgins and Lee, 1984)

5) The Longuet-Higgins and Lee metric, with Palmer and Krumhansl's "musician" weighting scheme (Longuet-Higgins and Lee, 1984; Smith and Honing 2006)

6) The Longuet-Higgins and Lee metric, with Palmer and Krumhansl's "non-musician" weighting scheme (ibid.)

7) Pressing's measure of cognitive complexity, which "distinguishes between five types of syncopation, and assigns weights to them according to their cognitive cost," which "depends on the placement of the rhythmic pattern within the metrical framework" (Gomez et al. 2007, 2)

8) The Toussaint measure, which "… exploits the metric hierarchy proposed by Lerdahl and Jackendoff (1984) [and] gives each unit of time a weight that depends on the relative strength of its metrical accent" (ibid.)

9) The rhythmic-oddity property, which measures the number of notes a rhythm has which partition the timeline into two equal halves. The greater the number of notes such as

these, the further a rhythm is from having the "rhythmic oddity property" (Gomez et al. 2007, 5)

10) The off-beatness measure, which in the music of Africa is simply the number of notes that do not fall on strong beats in a cyclical measure (ibid.)

11) Keith's measure of syncopation, which "relies on the definition of three musical events, namely, hesitation, anticipation, and syncopation, which are assigned weights according to their metric level" (Gomez et al. 2007,2)

12) The weighted note to beat distance measure, which "assigns weights to the onsets of a rhythmic pattern that depend on the distance (duration) of the onsets to the nearest strong beats of the meter" (ibid.)

13) The normalized pairwise-variability index, or nPVI, which compares changes in duration from one inter-onset interval to the next (Patel and Daniele 2003; Temperley 2017; Toussaint 2012, 2013)

14) Shannon entropy based on inter-onset intervals (Thul and Toussaint 2008)

15) Shannon entropy based on a binary stream of zeros and ones representing note onsets or lack of note onsets (deFleurian et al. 2014, 2017)

All performed well except the Tanguine measure, Lempel-Ziv complexity, the Longuet-Higgins and Lee measure with the non-musician weighting scheme (Palmer and Krumhansl 1990), Shannon entropy based on inter-onset intervals, and Shannon entropy based on a binary stream. The results regarding entropy will be discussed later.

**3.5 Markov Chains and Jazz Improvisation**

In the last two decades, many authors have explored computer generated jazz improvisation and automated musical analysis of jazz improvisation, frequently using Markov chains. Whereas in 1998, David Matthew Franz wrote that "mathematical investigations of a statistical nature regarding the analysis of improvisation are rare," by 2020 Tatsuyo Daikoku's paper "Statistical Properties in Jazz Improvisation Underline Individuality of Musical Representation" was not at all uncommon. The latter paper, in fact, is in a sense a precursor to the current work, distinguishing as it does between musicians (Bill Evans, McCoy Tyner, and Herbie Hancock) based on statistical properties of their solos.

Marom, whose dissertation "Improvising Jazz With Markov Chains" was published in 1997, was ahead of the curve. A number of papers came after.

A fascinating study by Gillick, Tang, and Keller (2010) uses clustering to construct probabilistic grammars of jazz, from which melodies are constructed using Markov chains of order one through four. Melodies were generated based on solos by trumpeters Clifford Brown, Miles Davis, and Freddie Hubbard, and listeners were asked to match originals to synthetically generated solos. Out of twenty listeners, 95% correctly identified Clifford Brown, 90% correctly identified Miles Davis, and 85% correctly identified Freddie Hubbard.

The "Jazzomat" project is a wide-ranging effort to explore systematically a number of theoretical aspects of jazz (e.g. Pfleiderer and Frieler, 2010). For present purposes, the interest of this project lies in its preliminary exploration of first-order Markov chains in analyzing one solo each by Dexter Gordon and Steve Coleman, reaching a number of *ad hoc* conclusions.

Chip Bell (2011) explores the combination of Markov techniques and "genetic algorithms" to generate compositions. "Genetic algorithms" in this context mean, very roughly, that users intervene in the automatic composition process (governed by Markov transition matrices) by selecting the best sounding compositions and allowing them to move on to the next stage of automatic composition.

According to Joseph Victor (2012), there are two goals that are frequently pursued in computational work applied to jazz: 1) to generate convincing jazz solos; and 2) given a jazz solo, to identify the underlying chord progression. He uses Markov chains to attempt the first task, and three techniques to accomplish the second: "a tree-metric-space model, a Naïve-Bayes model, and a Hidden Markov Model[4]." The first of these performed least well, while the third performed the best.

Whereas most Markov models of jazz improvisation are based on chord progressions, Norgaard, Montiel, and Spencer (2013) use a Markov model based solely on pitch to imitate the improvisations of Charlie Parker. Using a corpus of 48 solos by Parker to train their model, they found that it generated patterns similar to those found in the actual solos.

A straightforward application of Markov chains to the improvisation of music is pursued by Linskens and Schoenmakers (2014). According to them, the main contribution of their work is to study the transitions from composed melody to improvisation and back again.

---

[4] A "hidden-Markov model" is one in which a function of an underlying Markov chain is measured.

A surprising results is obtained by Frieler (2019), who explores solos in the Weimar Jazz Database using a number of quantitative parameters. Frieler tested the data to see whether or not improvised lines exhibit first-order Markovity. He finds that they do *not*; instead, he finds that the probability distributions describing those lines are *memoryless*. Given a field in which Markovity is generally assumed to hold, this result merits further testing.

Other papers on the intersection of Markov chains and jazz improvisation include Yun (2016), Rouse (2017), and Quick and Thomas (2019)[5,6].

## 3.6 Jazz Rhythm

The commonly held notion that jazz was a melding of European harmony with African rhythm is certainly an oversimplification, but it is a useful one nonetheless. According to Schuller (1968), syncopation was a translation of complex African rhythms onto simple European frameworks. This compromise led to ragtime, the craze of the 1890's (Blesh 2013); note that ragtime did not "swing" but was based on syncopations involving even sixteenth notes.

The time-feel called "swing" is characterized by uneven eighth note pairs, the first of each pair usually longer than the second (though see Iyer 2002 for a counterexample). In the hands of individual soloists, the exact ratios exemplify what Keil and Feld call "participatory discrepancies" (1994); more specifically, they are examples of what Iyer

---

[5] On the subject of stochastic music generation, the music of Xenakis and of Cage must be brought up, as these composers employed stochastic methods in constructing their works.
[6] Likewise, the Illiac Suite (1957) must be considered, as the first example of computer–based composition.

calls "expressive microtiming" (2002). The exact swing ratio employed by an individual musician is a musical parameter to be manipulated for expressive effect.

The complexity of African rhythm evolved atavistically through the simple syncopations of ragtime, the relative complexity of early jazz and the relative simplicity of swing into the extreme complexity of bebop. According to Jones:

*Bebop rhythm differs formally from swing rhythm, because it is more complex and places greater emphasis upon polyrhythmics.*

Jones 1963

Or again, according to Belgrad:

*Bebop's return to African-based polyrhythms was pioneered by drummer Kenny Clarke … Kenny Clarke led the way for Max Roach …*

Belgrad 1998

For present purposes, I will be interested primarily in two rhythmic criteria: 1) whether or not eighth-notes are swung, and; 2) whether or not the ratio of long/short swung eighth-notes is equal to 2:1.

# CHAPTER 4

# EXPERIMENT DESIGN AND RESULTS

## 4.1 Introduction

While experiments have been used to compare many computed measures of rhythmic complexity to perceived rhythmic complexity (see Chapter 3), to my knowledge, experiments directly involving entropy have been rare: an article by Thul and Toussaint (2008), and two by De Fleurian et al. (2014 and 2017). The first of these uses inter-onset intervals to calculate entropy and finds that "[t]he complexity measures based on statistical properties of the inter-onset interval histograms [including entropy] are poor predictors of syncopation or human performance complexity" (663). We will return to these conclusions later.

The articles by De Fleurian et al. represent rhythms using sequences of one's and zero's, with one's for note onsets and zero's for everything else. Thus the random variable at play has only two values. As discussed below, the methodology used here (and in Thul and Touassant), is completely different: the random variable in use represents inter-onset intervals between attacks, which can take on many different values depending on excerpt.

De Fleurian et al., in both of their papers, ask subjects to listen to test rhythms, and to decide whether they should be followed by a note or a rest, and to judge how easy it was to come to a decision. They average the results and correlate them with five computed information theoretic quantities: Shannon entropy (defined in Chapter 2),

entropy rate, excess entropy, transient information, and Kolmogorov complexity, and find that only entropy rate and Kolmogorov complexity predict experimental findings well.

The present work takes a more direct approach, by asking subjects to rank eighteen short rhythmic excerpts for complexity. It explores the effects of several variables, based on the hypothesis that they will prove to be significant factors in the perception of rhythmic complexity: entropy (calculated from IOIs between accented notes in jazz solos), periodicity, syncopation, number of notes, and jazz experience level. It takes into account order effects, and, finally, it explores how entropy and the other variables compare and contrast with one another.

It will be important in what follows to distinguish between these related numerical concepts. In particular, while in this chapter, perceived complexity is the dependent variable of interest, in the following chapter, entropy is the dependent variable of interest. Note that *all* notes in each experimental excerpt are counted in this chapter, while in the following chapter using transcriptions, only *accented* notes are counted. Experimental excerpts were constructed from transcriptions by isolating accented notes.

For the most part, periodicity and syncopation depend on meter while entropy does not, for the following reasons. Strong periodicity, defined here as overlap at a distance that is a multiple of four eighth-notes, implies a strong relationship to the underlying quadruple meter. And syncopation is defined explicitly in relationship to the underlying meter. Meanwhile, an excerpt can exhibit high periodicity but low entropy: an example would be a perfectly periodic rhythm with only one or a few IOIs. And shifting a rhythm by one eighth-note would completely change syncopation without changing

entropy at all. I will adopt a sensitive statistical approach that will allow me to determine the relative role each of these factors.

## 4.2 Experiment Design

Fifteen music majors at UMass Amherst were asked to listen to eighteen short eighth-note-based rhythmic excerpts and to rate them for "complexity" on an integer scale from 1 to 7. A strict definition of "complexity" was not given; subjects were relied upon to provide their own, subjective, definition. They were asked to listen to each excerpt twice before making a judgement.

The excerpts used for this experiment were derived by selecting excerpts from solos by Armstrong, Hawkins, Young, Christian, and Parker, and isolating only dynamically accented notes. This is an example of "reduction," which has a long history in music theory[7]

Pfordresher (2003) demonstrates that listeners are indeed sensitive to reduction, in this case derived from melodic (contour) and rhythmic (agogic or durational) accents. He proves this by asking listeners to tap along with test rhythms. Pfordresher and Jones (1997) study the ability of listeners to tap along with rhythms having both concordant and discordant "Joint Accent Structures" (JASs), that is, with rhythms having simple and complex temporal relationships between accent types. They found that listeners had an easier time tapping along with rhythms having concordant JASs, further evidence that

---

[7] Riepel was well-known for analyzing music in terms of phrases, a form of reduction (Christensen 2010, 671). Schenker is perhaps the best known proponent of reduction, showing how any tonal composition can be reduced to an *Ursatz*, consisting of a $\hat{1} - \hat{5} - \hat{1}$ bass motion (*Bassbrechung*) and a three, five, or eight note descending melodic line (*Urlinie*). Schoenberg was best known for his use of twelve tone rows, but in *Harmonielehre* also addressed motives, phrases, and themes (Krämer 1993). And Robert Morris, in a 1993 paper, introduced the *contour reduction algorithm.*

listeners are sensitive to reduction. Ellis and Jones (2009) showed that listeners had an easier time tapping along to test rhythms when accents were made more salient.

Lehrdahl and Jackendoff (1983) propose a system for reducing musical surfaces using "time span" and "prolongational" techniques. The phenomonal accents – one of three types of accents they propose (17) – in common practice music, however, usually create or reinforce metric or structural accents. This is not the case in syncopated music such as jazz or rock. Temperley (1999) addresses this issue by proposing a "deep structure" that in essence "un-syncopates" phenomenal accents, making them fit in with Lehrdahl and Jackendoff's model. Temperley points out that when whistling or humming along with a syncopated melody, people often reproduce the "deep structure" of a tune rather than the actual tune, again demonstrating the perceptibility of reduction.

Additionally, with regards to the question of whether or not listeners hear reduced versions of musical surfaces as having come from the surfaces being reduced, several authors have experimentally tested this, and have found an affirmative answer. These author include Dibben (1994), Carrabré (2015), and Yust (2012).

The excerpts used in this experiment consisted of 20 measures of 4/4 time. In addition to the rhythms of interest, a quarter-note click track emphasizing the first beat of every measure was used. The click track ran for all 20 measures, while most rhythms of interest began after four pickup measures. Some rhythms began after just two or three pickup measures, while some began after five. Most rhythms ended in the 20th measure, but some ended earlier. An example stimulus excerpt is shown in Figure 13; all test rhythms are provided in Appendix A.

**Figure 13.** Example of stimulus excerpt

Recall that in Chapter 2 I presented an example of how to obtain a probability distribution from a simple musical excerpt. Once a probability distribution has been obtained, entropy can be calculated using Equation 2.1. In Chapter 2, the distribution was shown to be as follows in Table 4.

| IOI (eighth-notes) | 2 | 6.66 | 7.33 | 10 |
|---|---|---|---|---|
| Probability | 1/5 | 1/5 | 2/5 | 1/5 |

**Table 4.** Distribution for excerpt from Chapter 2 (same as Table 3)

From this distribution, entropy can be calculated directly from Equation 2.1. That equation is given here for ease of reference:

$$H \equiv - \sum_x p(x) \log(p(x)) \qquad \text{Eq. 2.1}$$

The calculation is then given by:

$$H = -(1/5)\log_2(1/5) - (1/5)\log_2(1/5) - (2/5)\log_2(2/5) - (1/5)\log_2(1/5) = \quad \text{Eq. 4.1}$$

$$(3/5){*}\log_2(5) + (2/5)\log_2(5/2) = 1.922$$

While this is an artificially simple example, it captures the essence of how I calculate entropy for musical excerpts. Note, by the way, that the entropy calculation does not explicitly rely on the values of the IOIs; all that matters is the form of the probability distribution itself.

Only notes and rests evenly divisible by eighth-notes were used. The reason for this was that including other rhythms, such as triplet- or sixteenth-note-based rhythms, or rhythms involving ornamental straight-eighth-notes, would have required a longer experiment run time, which was already about twenty minutes. Furthermore, as will be described below, restricting the rhythms to eight-note multiples enables us to quantify syncopation more simply. It will be assumed that the conclusions obtained here, using excerpts divisible by eighth-notes, are generalizable to a broader range of rhythms.

Excerpts were generated using Finale. The woodblock sound on E5 was used for the rhythm of interest, and the clave sound was used to generate the quarter-note click track. Both of these sounds have sharp percussive attacks and no sustain, so that, for example, an eight-note followed by an eighth rest sounded exactly the same as a quarter-note. Excerpts were played at 150 bpm, and swing eighth-notes were generated using the Finale playback option "swing = 50," corresponding to quarter-note triplet-eighth-note triplet pairs.

Entropy was the primary variable of interest in this study. In order to study the relationship of entropy to perceived complexity, however, it was necessary to consider

other predictor variables as well. First and foremost, it was necessary to control for number of notes.

Since the excerpts were roughly the same number of measures ($\mu$ = 16.06, $\sigma$ = 1.21, min = 13, max = 18), controlling for number of notes was similar to controlling for *density* of notes; it is not possible to decide *a priori* which quantity a listener would perceive. The importance of this dichotomy will be returned to in Chapter 5, in the context of discussing the commonly held notion that Charlie Parker's solos are more unpredictable than Coleman Hawkins's. The variation in excerpt length was the consequence of selecting excerpts from a collection of solos and selecting mainly 16-bar excerpts with one or possibly two pickup measures and sometimes ending early.

Three ranges of entropy were crossed with three ranges of number (or density) of notes and for each of eight combinations, two excerpts were selected. One excerpt, however (high entropy, low number of notes), was also used for the combination high entropy, medium number of notes.[8] Table 5 shows the ranges of entropy and number of notes.

| Variables/Ranges | *Low* | *Medium* | *High* |
|---|---|---|---|
| *Entropy* | 1.7 – 2.3 | 2.3 – 2.9 | 2.9 – 3.5 |
| *Number of Accents* | 16 – 23 | 24 – 30 | 31 – 38 |

**Table 5.** Ranges of entropy and number of notes used for selecting excerpts

Note that for computational purposes, variables were treated as continuous; the ranges referred to here were used only for selecting excerpts. For some excerpts, entropy

---

[8] This was inadvertent, the result of translating an experiment that was intended to be conducted in person to a virtual platform. Fortunately, due to the built-in redundancy of the experiment, this did not jeopardize the results. In fact, the duplicate excerpt facilitated the deduction of the form of the term for carryover effects (Anna Liu, personal communication).

values were on the borderline between two ranges; this did not present a serious problem, since analysis used actual entropy values rather than entropy range

Next it was necessary to control for the interaction between rhythm and meter. This was done in two ways: by including periodicity as a quantity of interest, and by treating syncopation as a quantity of interest.

To control for *periodicity* of the test rhythms, or lack thereof, I employed three autocorrelation-like variables I will call *corr.4*, *corr.8*, and *corr.16*. They represent the number of inter-onset intervals (not necessarily adjacent) of duration four, eight, or sixteen eighth-notes long. If there is a good deal of correlation at a distance of four, eight, or sixteen eighth notes, these variables will indicate so. And if the test rhythms are strongly tied to the underlying meter, that will be reflected in the values of these variables. The formulas are presented below, where *s* is the "signal" (0 or 1) corresponding to a particular excerpt, indexed by eighth-note position in an excerpt *n* eighth-notes long[9]:

$$corr.4 = \sum_{i=5}^{n} s(i)s(i-4)/n \qquad\qquad \text{Eq. 4.2(a)}$$

$$corr.8 = \sum_{i=9}^{n} s(i)s(i-8)/n \qquad\qquad \text{Eq. 4.2(b)}$$

$$corr.16 = \sum_{i=17}^{n} s(i)s(i-16)/n \qquad\qquad \text{Eq. 4.2(c)}$$

Each of these sums reflect the overlap between the signal and a shifted version of the signal. For example, if we shift the signal by four eighth-notes, we can only evaluate the sum from the fifth member of the signal to its end, because the *s(i–4)* term precludes using anything before the fifth member of the signal. The last term in the sum will be *s(n)\*s(n–4)*. No further terms enter into the sum because the signal has only *n* values.

---

[9] Results were also calculated using $n^2$ in the denominator instead of *n*. They were essentially unchanged.

I also evaluated the syncopation within each excerpt. Syncopation and entropy are two conceptually distinct quantities, both of which might contribute to perceived rhythmic complexity. In particular, entropy is independent of meter, while syncopation depends on it strongly. To control for syncopation, I employed the metric of Longuet-Higgins and Lee (1984), introduced in Chapter 3, divided by number of notes, to yield a variable I will call "LHL quotient."

While Smith and Honing (2006) and Fitch and Rosenfeld (2007) have demonstrated the LHL metric to be an empirically meaningful quantity, the fact that a single sounded note can participate in multiple syncopations would seem to be counter-intuitive (see Section 3.3), particularly when a sounded note is followed by one or more measures of rest. In such a situation, a single note – say on the and of four – can be followed by syncopation totals of 7 (a large number in this context)[10] for each measure of rest following the sounded note. In the excerpts used here, there are frequently one or two measures of rest, while in the excerpts used by Smith and Honing (who use the data of Shmulevich and Povel, 2000) and by Fitch and Rosenfeld, there do not appear to be any measures of rest. This should be addressed in a later study.

Next, carryover effects were taken into account. (Note that eight different excerpt orderings were used). It was found that a relatively high complexity rating exerted an upward pressure on the following rating, while a relatively low complexity rating exerted a downward pressure on the following rating. This effect was significant in all models. It was, however, a small effect, and did not qualitatively affect the outcomes of the experiment.

---

[10] This assumes that only eighth-note subdivisions are used, not sixteenth-note subdivions. Includingsixteenth-notes would not assuage the problem; in fact, it would make it worse.

One final factor was taken into account: the jazz experience level of the participants. This was recorded as a binary variable indicating whether or not the subject was a jazz musician (undergraduate music majors at UMass Amherst study either jazz or classical music). It did not appear to make any difference in the analysis, and was therefore ignored.

.

## 4.3 Experimental Results

The experimental data is shown in Figures 14 and 15. In Figure 14, entropy is plotted on the x axis and complexity ratings averaged over participants are plotted on the y axis.. In Figure 15, number of notes is plotted on the x axis and complexity ratings averaged over participants are plotted on the y axis. In both graphs there is a general upward trend. Error bars indicate standard error.
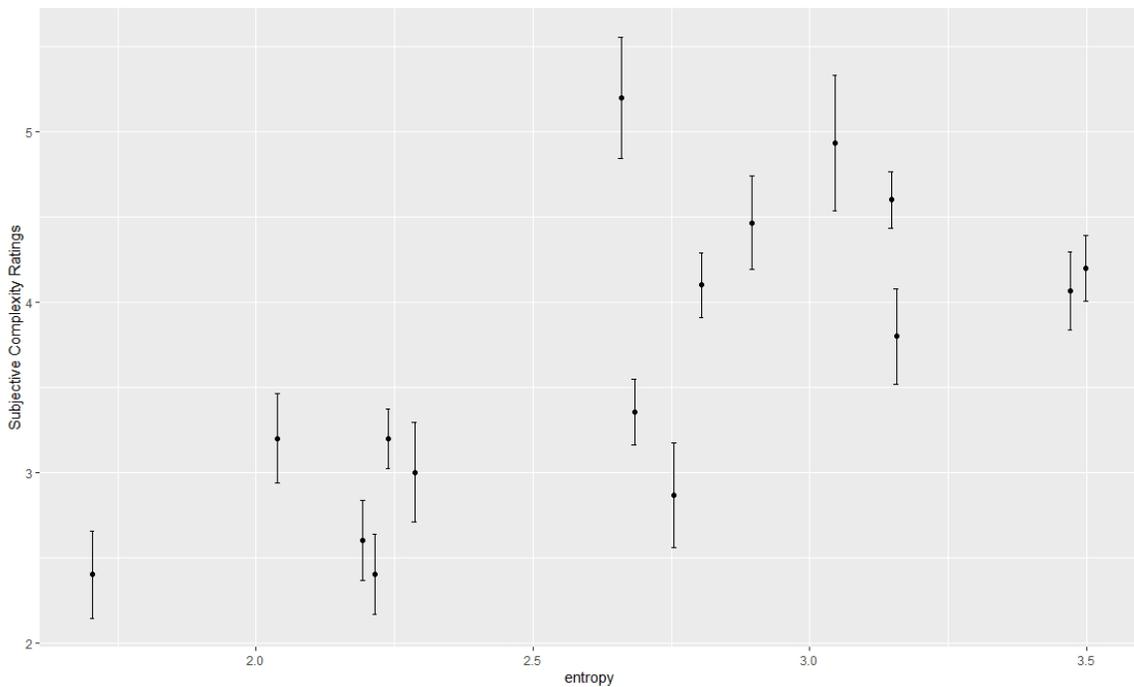


**Figure 14.** Experimental data. Entropy is plotted on the x axis and complexity ratings averaged over participants are plotted on the y axis. With standard error.
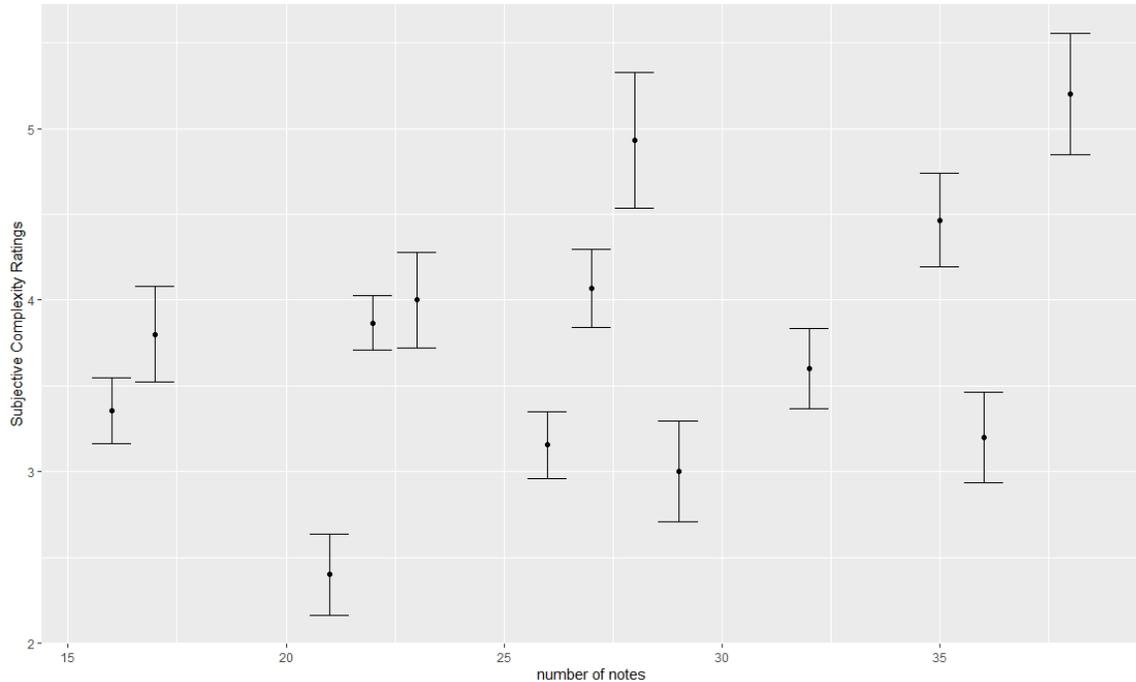
**Figure 15.** Experimental data. Number of notes is plotted on the x axis and complexity ratings averaged over participants are plotted on the y axis. With standard error.

Data were analyzed using a mixed effects multivariate regression model[11]; the Satterthwaite approximation for effective degrees of freedom was used since the variance of the data was not known a priori. A "mixed effects" model was necessary because some factors were "fixed effects" (entropy, number of notes, periodicity, LHLQ, part of carryover effects), while others were "random effects" (participant, excerpt, part of carryover effects). In general, a fixed effect is one whose specific values we are interested in, while a random effect is one that is chosen merely to be a representative of a larger population of values. Data were analyzed treating all factors of interest as continuous variables (this does not include experience level), and also treating the perceived complexity ratings as a continuous variable.[12]

---

[11] On the advice of Anna Liu, personal communication

[12] The question of whether or not Likert–scale variables – such as the response variable in this experiment – can be treated as continuous variables is hotly debated, as a quick google search will indicate. There seems

There are four assumptions that must ordinarily be met in order to employ multiple regression. They are: 1) independence, 2) normality, 3) homoscedasticity, and 4) linearity. For this data, the first assumption is not met, since order effects are significant (each complexity rating depends on the previous one). However, in a personal communication, Anna Liu stated that this model was valid despite the lack of independence, because a random effects term for carry-over effects was included in the model. Normality, the second assumption, was tested using the Shapiro-Wilks test on the residuals of each complete model discussed below: for the models that will be labelled "short" and "long," the $p$ values were 0.5727 and 0.3897 respectively. Homoscedasticity was tested by plotting residuals against each predictor variable, and noting that the spread of the residuals did not depend too drastically on any of the predictor variables. Linearity is confirmed by examining the dependence of residuals on each predictor variable; no significant structure was found, indicating that linearity holds (Michael Lavine corroborated this result).

As a first step in analyzing the data, a backwards selection algorithm was used. This algorithm starts with a full model, and progresses in a step-wise fashion by deleting one variable at a time according to which change will yield the greatest reduction in the Akaiki Information Criterion until some prespecified stopping condition is reached (all of this was done using the 'step' command in the programming language R). The backward selection algorithm does not necessarily yield the absolute *best* model, but is merely a heuristic to identify a *plausible* model for further consideration.

---

to be a consensus that a minimum of seven response choices (such as used in this experiment) is required for such a variable to be treated as continuous. However, personal communications from Anna Liu indicate that it is probably fine to treat the Likert–scale data as continuous for this experiment. This is a common assumption in the literature, and it is the assumption I adopt here.

The Akaiki Information Criterion, and its relative the Bayesian Information Criterion, are both derived from the log likelihood function. The likelihood function treats observed variables as parameters and fit parameters as variables, and is maximized for a maximum-likelihood fit. Using the logarithm of the likelihood function is more convenient than using the likelihood function itself due to the very small numbers involved.

Both the AIC and the BIC yield smaller values for better models. Both penalize a large number of parameters and favor simpler models. The AIC tends to be more liberal in terms of including parameters, while the BIC is more conservative. The Akaiki Information Criterion is defined by:

$$AIC = 2k - 2ln\hat{L} \qquad\qquad \text{Eq. 4.3}$$

where $k$ is the number of parameters in the model and $\hat{L}$ is the maximized likelihood function. The Bayesian Information Criterion is defined by:

$$BIC = k\ ln(n) - 2ln\hat{L} \qquad\qquad \text{Eq. 4.4}$$

where $n$ is the sample size.

The typical mixed-effects multiple regression fit in the statistics programming language R uses the Restricted Maximum Likelihood (REML) of a model for a set of data. However, fitting data this way precludes the use of AIC to compare models exactly, including in the process of backwards selection, since the AIC is based upon maximizing the likelihood function rather than the "restricted" maximum likelihood. Furthermore, it complicates the process of testing for interaction effects. Therefore, I opt for a "maximum likelihood" method rather than the "restricted maximum likelihood" method here during model identification and interaction testing, then switch back to REML for

reporting the final results. Choosing REML methods over maximum-likelihood methods does not qualitatively affect final *p*-values (in other words, significance of variables or lack thereof does not change).

With this in mind, I return to the backward selection algorithm. In the present context, it identifies as a good model one that includes: *number of notes*, *corr.4*, *LHLQ*, order effects, and a random effect for excerpts as parameters. The coefficient corresponding to *corr.4* is negative, indicating that perceived complexity has an inverse relationship to *corr.4*: decreasing periodicity increases complexity rating. The other coefficients are positive. This model will be called the long model.

Since I am primarily interested in entropy, I also propose a model using *number of notes*, *entropy*, order effects, and a random effect for excerpts. This will be called the short model.

Both the short model and the long model were tested for interactions. Interactions occur when the effect of one variable depends on the value of another. There are multiple ways of testing for interactions; I used two. For both methods, I examined all permutations of fixed effect terms concatenated by either addition or addition and interaction.

The first method I employed was simply to examine the term-by-term significance of the interactions between predictor variables. This method revealed no significant interactions (i.e. no *p*-values less than 0.05).

The second method of testing for interactions again uses models with terms concatenated by addition or addition and interaction. As opposed to the previous method, however, this method penalizes the increase in complexity resulting from the addition of

interaction terms. It is based on the concept of the ANOVA procedure, but compares two models rather than calculating a single model. In the R programming language, the syntax is simple: anova(model1,model2). *P* values less than 0.05 indicate that adding one or more interaction terms results in a favorable model. Using this method did not reveal any *p* values less than 0.05, thus, suggesting, too, that interactions be ignored.

Table 6 shows the AIC and BIC for the long model and the short model, as well as the $R^2$ metric for both models[13]; lower values of the AIC and BIC indicate better fit, while higher values of $R^2$ indicate a better fit. The rule of thumb for the AIC and BIC is that a difference of 2.0 or less is insignificant; differences of 10.0 or more are extreme. As Table 6 shows, the long model is better than the short model according to the AIC, and *slightly* better than the short model according to the BIC. This is consistent with the description of the two metrics above as more liberal (AIC) or more conservative (BIC), since the long model has more parameters than the short model. Table 6 also shows that the short model is slightly better than the long model according to the $R^2$ metric. $R^2$ values in the 0.5–0.6 range indicate robust models, particularly for studies involving human psychology and performance.

| Model/Criterion | AIC | BIC | $R^2$ |
|---|---|---|---|
| *Long Model* | 708.214 | 743.548 | 0.553 |
| *Short Model* | 715.181 | 746.981 | 0.570 |

**Table 6.** Goodness of fit comparisons using Akaike Information Criterion, Bayes Information Criterion, and $R^2$

---

[13] The Nakagawa $R^2$ for linear mixed models was used.

*P*-values for the long model are shown in Table 7, and reveal that all included

parameters are significant. This is not surprising, as we expect syncopation and

periodicity to be reflected in complexity ratings, in addition to number of notes. The *p*-

values for the short model are shown in Table 8, and reveal that all included parameters

are significant. This, too, is to be expected, as we expect entropy as well as number of

notes to be significant. The conclusion is that the models based on lack of

periodicity/syncopation and on entropy are both reasonable models in terms of goodness-

of-fit and model complexity.

| Variable | *p* Value |
|---|---|
| Number of notes | 0.000158 |
| *Corr.4* | 0.0000382 |
| LHLQ | 0.012290 |
| Order effect | 0.00694 |

**Table 7.** Description of long model

| Variable | *p* Value |
|---|---|
| Number of notes | 0.01624 |
| Entropy | 0.000215 |
| Order effect | 0.008324 |

**Table 8.** Description of short model

An issue that can adversely affect the interpretation of multiple regression results

is "collinearity" or "multicollinearity". This is a situation that obtains when one or more

predictor variables can accurately be derived from one or more other predictor variables by linear combination. This is undesirable since in this situation, the coefficient estimates of the multiple regression may change erratically in response to small changes in the model or the data (*Wikipedia*). An easy way to test for multicollinearity is to use the so-called *variance inflation factor*. This represents the ratio of the variance as it occurs for each term in a model to what that variance would be in the case of perfect independence between variables. VIF's greater than 2.5, 5.0, or 10.0 – depending on the analyst's choice – reveal the presence of multicollinearity. Applying this technique to the model at hand, using thresholds as low as 2.5, revealed that multicollinearity is not a problem.


## 4.4 Discussion

While it may appear that the models presented above are independent of one another, they may in fact be an example of "statistical mediation." Statistical mediation happens when a random variable "A" influences a random variable "C" through a "mediating variable" B. In the current situation, A could be "entropy," C could be "perceived rhythmic complexity," and B could be "*corr.4*." This is supported by the fact that there is a strong negative correlation between entropy and *corr.4* (–0.80) and the fact that the backwards selection algorithm selected "*corr.4*" but not "entropy." What this means is that for excerpts with a high value of *corr.4* there are many IOIs of four eighth-notes, and thus a low entropy, while for excerpts with a low value of *corr.4*, there is a more diverse distribution of IOIs and thus a higher entropy. So it appears that this is a

case of statistical mediation, without changing the fact that entropy appears to result in higher subjective complexity ratings.

Overall, then, the experiment revealed that for rhythms comprised solely of eighth-notes and rests and their integer multiples, entropy differences are reflected in subjective rhythmic complexity ratings (though this is probably mediated by periodicity); that periodicity or the lack thereof and syncopation are reflected in complexity ratings; that the number of notes (when excerpt length is held roughly constant) is reflected in complexity ratings; and that carryover effects are reflected in complexity ratings. More work on the effects of jazz experience level may be called for.

These results challenge the finding of Thul and Toussaint (2008) that findings based on the statistical properties of IOI histograms (e.g. entropy) do not predict syncopation or human performance metrics. There could be several reasons for this. First of all, as previously discussed, syncopation and entropy are two completely different things; thus the finding that entropy does not predict syncopation is not at all surprising. Perhaps the null result regarding human performance metrics reflects a fundamental difference between measuring human performance metrics and measuring subjective complexity ratings. Perhaps the null result discrepancy comes from the fact that Thul and Toussaint used sixteenth-note subdivisions in addition to eight-note subdivisions; this methodology may or may not be more accurate than using just eighth-note subdivisions. Or perhaps these are merely results that disagree with each other, calling for more work to determine which is correct. In any case, further studies are called for.

In a future study, it would be interesting to study the carryover effects of *actual* independent variables (entropy, number of notes, *corr.4*, *corr.8*, *corr.16*, LHLQ). In this

scenario one might expect a high entropy value, for example, to exert a downward pressure on the following rating, while a low entropy value might exert an upward pressure on the following rating. In other words, having just heard a highly entropic excerpt, for example, a subject might hear the following excerpt as being *less* entropic by comparison.

**4.5 Conclusion**

In this experiment, fifteen music majors rated eighteen rhythmic excerpts (seventeen with one duplicate) for complexity. Two models were used to describe the data well ($R^2 \approx 0.6$): one including entropy (calculated from inter-onset intervals between notes in each excerpt) and number of notes, and one including number of notes, periodicity, and syncopation. It is possible that the two models are an example of statistical mediation, in which entropy influences periodicity (or lack thereof), and periodicity in turn influences perceived complexity.

In this experiment, only rhythms divisible by eighth-notes were used. This was necessary because including a wider array of rhythmic subdivisions would have made the experiment's run time too long. Another experiment including a wider array of rhythmic subdivisions will have to await another study. The results presented here, however, already suggest strongly that perceived rhythmic complexity depends on entropy, when entropy is calculated using probability distributions obtained from measuring IOI's between dynamically accented notes in jazz solos. In the following chapters, it will be assumed that entropy is, indeed, perceptible. If it is not, however, it is interesting in its own right, as a signature of musical style. This will be demonstrated in the next chapter.

# CHAPTER 5

# COMPUTATIONAL RESULTS

## 5.1 Introduction

My main goal in this chapter is to understand whether entropy, on the whole, is a signature of musical style among the five musicians considered in this study. This is an interesting question because the results from Chapter 4 seem to indicate that entropy is an indicator of rhythmic complexity, though the transcribed solos analyzed here frequently contained a wider range of rhythmic subdivisions than those contained in the experimental excerpts. To be specific, they contained eighth-note and quarter-note triplets, sixteenth-notes, and ornamental straight-eighth-notes. Even if entropy is *not* an indicator of rhythmic complexity for this wider range of rhythmic subdivisions, however, the question of whether or not entropy is a signature of style is still an interesting one, if for no other reason than academic curiosity.

## 5.2 Corpus of Transcribed Solos

The computations described here are based upon a corpus of 88 transcriptions of solos by five great jazz musicians: Louis Armstrong (1901–1971), Coleman Hawkins (1904–1969), Lester Young (1909–1959), Charlie Christian (1916–1942), and Charlie Parker (1920–1955). Note that the birthdates of these musicians span the first two decades of the twentieth century. The transcriptions were done by the author, using the Omnibook as a starting point for the Parker transcriptions.

Transcribed solos were converted to Excel files by evaluating the elapsed times, in eighth notes, from the beginnings of solos to the onset of accented notes. For some

double-time solos, e.g. on ballads, elapsed times were sometimes evaluated in sixteenth notes. Eighth-notes (or sixteenth-notes on some double-time solos) were assumed to be swung unless they were flagged as "straight."

Efforts were made to select solos representative of each soloist's different style periods, and to include chronologically even representations of each soloist's oeuvre. For Armstrong, solos were included from his Hot Fives and Hot Sevens, from Louis Armstrong and His Savoy Ballroom Five, from Louis Armstrong and His Orchestra, and from Louis Armstrong and His All-Stars. (His early work with King Oliver was not included). Representatives of Armstrong's trumpet solos and of his vocal work were included. For Hawkins, his work with Fletcher Henderson was represented, as was his famous recording of "Body and Soul" from 1939 (his work in Europe was not included), his tenure as a leader on 52$^{nd}$ St., his work with Thelonious Monk, and his later work with pianists Tommy Flanagan and Paul Bley. Lester Young's work with Basie was represented, as was his work with Billie Holiday, his work as a leader both before and after his enlistment in the Army, and his late work with Jazz at the Philharmonic. Charlie Christian's premature death resulted in a lack of discernible style periods; the years 1939–41 were covered roughly equally. Finally, the solos of Parker, according to Kernfeld (1996), can be divided into his early style (pre-1944) and his mature style (post-1943). Both periods are represented here.

All excerpts were in 4/4 time, sometimes articulated as 12/8 time, with a tempo range of 63 (Armstrong "What a Wonderful world") to 280 bpm (Parker "Honeysuckle Rose" and "Crazeology").

The smallest subdivision used in the transcriptions was the sixteenth note (or in the case of double time solos, the thirty-second note). This subdivision was sufficient for identifying accents in the solos transcribed. Efforts were made to accurately represent rhythmic anticipations and suspensions that approximate eighth-note rhythms, for example in Armstrong's solos on "Stardust" or "What a Wonderful World." Straight eighths were distinguished from swing eighths, and in some solos based on double time (e.g. Hawkins, "Wanderlust"), sixteenth notes were treated as swung notes, with some straight sixteenths.

The shortest excerpt was 12 bars with pick-ups (Parker, "Hootie Blues"), while the longest was 200 bars (Young, "Ad Lib Blues"). The question of minimum solo length is an important one, since as the number of bars tends toward zero, so must the excerpt's entropy. For the time being, 12 bars were considered long enough to produce a meaningful entropy value. As will be discussed later, however, solo length will be accounted for by treating the number of accents as a covariate in the analysis of entropy's dependence on musician.

Only dynamically accented notes were used in the calculations. Accented notes were used because they reflect a layer of structure superimposed by the soloist on the structure of the solo. Dynamically accented notes were used on the assumption that they are the easiest to perceive. Later in the study, contour accents will be investigated, and in a future study, other kinds of accent might be included.

Joel Lester, quoted in John Roeder's "A Calculus of Accent," defines accent in terms of the music surrounding a given accent, specifically "the relative strength of a note or other musical event in relation to surrounding notes or events." An archetypical series

of accents, from Lester Young's solo on "Blues For Greasy," is shown in Figure 16 (this was seen earlier in Figure 1). A more complex example, from Coleman Hawkins's solo on "Body and Soul," is shown in Figure 17; here a phrase played *mezzo piano* is followed by a phrase played *mezzo forte* containing an accent on the second beat of the second phrase. The first note of the second phrase is *not* considered to be an accent, even though it is louder than the preceding phrase. Two examples from the Louis Armstrong trumpet corpus illustrate how accents interact with register; this is of interest because of the strong dependence of loudness on register in trumpet music. In the first example, from his solo on *Stardust*, an ascending phrase lands on an accented note which is then repeated with accents (Figure 18). I perceive these notes as being accented even though there is a crescendo leading up to them. An example from "I Double Dare You," by contrast, is in a moderately high register, but does not accent the highest note in the phrase (Figure 19). Finally, I discuss one of Armstrong's vocal solos on "What a Wonderful World" (see Figure 45, later in the document). On the bridge, there are many repeated accented notes, but there are enough notes in the solo that are *not* accented to make the identification of accented notes meaningful.



**Figure 16.** A straightforward example of accents from Young, "Blues For Greasy"



**Figure 17.** An excerpt from Hawkins's "Body and Soul" in which a change in dynamic levels does not cause an accent

**Figure 18.** An excerpt from Armstrong's "Stardust" in which a crescendo does not negate the perception of accents



**Figure 19.** An excerpt from Armstrong's "I Double Dare You" in which register does not solely determine accents

Most of the eighth note pairs considered here are swing eighths, meaning that they can be grouped into long-short pairs starting on quarter-note beats. According to Collier and Collier (2002), for example, Armstrong sometimes employs eighth note pairs with a durational ratio of 1.6:1 (or 8:5). In many excerpts, however, Armstrong and other soloists employ a ratio of 2:1.

A ratio of 2:1 creates a triplet feel, since it corresponds to a single quarter-note triplet followed by a single eighth-note triplet. This ratio implies a compound 12/8 meter, and the rhythm section often supports this; a prime example is one of Armstrong's performances of "What a Wonderful World," from 1967, transcribed in Figure 45. The piano plays chords in 12/8 time, and this provides the backbone upon which Armstrong's solo is built.

For some excerpts, it must be specified whether the swing ratio is 2:1 or not in order to calculate entropy from the transcribed solos, since if it *is*, the IOIs involving swing eighth notes are equivalent to rhythms involving triplets, thus potentially reducing the total number of distinct IOIs in the excerpt. This causes a potential problem since

75

there is theoretically a discontinuity in entropy as a function of swing ratio: as 2:1 is approached, the entropy is calculated according to the formula for *non*-triplet swing eighth notes, but this jumps to the *triplet* version of the entropy when 2:1 is reached. For other excerpts, this does not matter because there are no accents on the third triplet of any quarter-note beat.

For ratios other than 2:1, the exact ratio is unimportant; what matters is that the difference between groups of eighth note beats starting on a "long" eighth note rather than a "short" eighth note is taken into account in the analysis. This is true for groups consisting of odd numbers of eighth notes. For example, a group of three eighth notes starting on a long eighth note beat would have a duration of L+S+L, while the duration of three eight notes starting on a short eighth note beat would be S+L+S, where S and L stand for the (unequal) durations of the short and long eighth note beats respectively. No soloist employed dotted-eighth/sixteenth swing eighth notes, which would potentially create difficulties distinguishing between swing eighth notes and sixteenth notes.

To facilitate analysis, only excerpts in which the swing ratio is obviously 2:1 or in which the entropy does not depend on the swing ratio were included in the corpus. This excludes excerpts in which the swing ratio is difficult to ascertain *and* in which the entropy depends on whether or not the swing ratio is 2:1.

One solo, it should be noted, contained eighth-notes that are almost even, or "straight" in jazz parlance. In particular, Charlie Parker's solo on "KoKo" is built using eighth-notes that are either straight or very close to being straight. Solos such as these cannot be included in the corpus, because the decision of whether to treat them as swing

or straight in calculating entropy has a marked effect on the outcome (on the order of 10% for "KoKo").

For a chart containing musician names, excerpt names, recording dates, a binary variable called "TS" for "triplet swing," which indicates whether or not an excerpt exhibited a 2:1 swing ratio AND the entropy depended on whether or not the swing ratio was treated as 2:1, number of distinct IOIs, number of accents, and entropy, see Appendix B at the end of this paper.

## 5.3 Methods

The primary tool employed for the task of determining whether or not entropy depends on musician is that of *estimated marginal means*, or EMMs. EMMs allow us to study the dependence of a continuous "response" variable (in this case, entropy), on a single discrete "factor" (in this case, musician), in the presence of continuous "covariates" which may depend on the factor. *P* values for each pairwise combination of musicians reveal whether or not entropy depends on musician.

As covariates to include in this model, I selected number of distinct IOIs and number of accents. Not only are these intuitive choices – since both of them could be reflected in entropy values and both of them could depend on musician in the corpus used here – they also sidestep the issue of normalization. Rather than normalize by maximum entropy for a given number of IOIs, I use number of IOIs as a covariate. This is also a convenient way to sidestep the issue of sample size.

To compare different musicians in terms of entropy, the present study takes a different approach from that used in previous studies (Youngblood 1958; Knopoff and

Hutchinson 1981, 1983; Snyder 1990): rather than grouping data together by composer, different excerpts are evaluated for entropy separately, and the estimated marginal means technique lets us evaluate differences in entropy for significance between musicians.

In order to justify the inclusion of number of distinct IOI's and number of accents as covariates, I turn to a graphical technique called the "added variable plot." Added variable plots reveal whether or not it makes sense to add a variable to a linear model. First, one constructs two models: $y \sim x_1 + x_2 + \ldots + x_n$ and $x_{n+1} \sim x_1 + x_2 + \ldots + x_n$, where $y$ is the dependent variable, $x_1 \ldots x_n$ are the independent variables already included in the model, and $x_{n+1}$ is the independent variable to be tested for addition. One then calculates the residuals of the two models and plots them against each other. A straight line indicates that the variable $x_{n+1}$ should be added to the model.

To see this, consider the case of adding a second independent variable to a simple linear regression plot (Weisberg 1980, 35–39). We are interested in the part of $y$ that is not described by $x_1$ (thus we look at the residuals of $y \sim x_1$) and how that depends on the part of $x_2$ that is not described by $x_1$ (thus we look at the residuals of $x_2 \sim x_1$). A straight line indicates that $y$ depends linearly on $x_2$ in the context of a multiple regression model with $x_1$ and $x_2$ as independent variables.

In this case, I begin by considering the addition of "number of distinct IOIs" to the single variable "musician"; the corresponding plot, Figure 20, shows that the variable *should* be added. The same procedure is used to show that "number of accents" should be added to "musician" and "number of distinct IOIs" (Figure 21). According to statistician Michael Lavine (personal communication), these are "textbook examples."

Having selected number of distinct IOIs and number of accents as covariates based on our intuition about the factors (other than musician) upon which entropy should depend, and based on the added variable plots, I now test the resulting model to see if it meets the formal requirements for multiple regression, the precursor to using the EMM procedure.
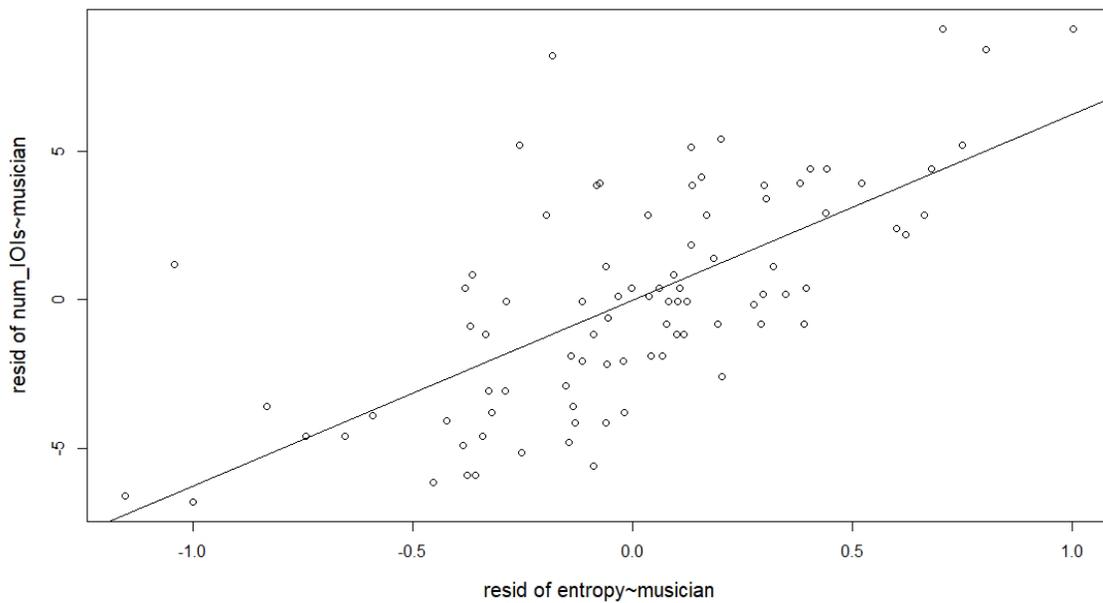


**Figure 20.** Added Variable Plot for number of IOIs

The four requirements for multiple regression, as discussed in Chapter 4, are: 1) independence, 2) normality, 3) homoscedasticity, and 4) linearity. Independence follows from the fact that the entropy values do not influence one other. Normality can be tested by calculating residuals from the multiple regression model *entropy ~ musician + num_IOIs + num_accents* and testing them for normality using the Shapiro-Wilks test: $p = 0.120$. Homoscedasticity can be tested using a plot of residuals vs. fitted values; random scatter plots indicate that variances are equal (Figure 22). This type of plot will be used again later. Other plots, not examined here, can also be used to test for

homoscedasticity[14]. Finally, the fact that the residuals do not show any systematic

variation when plotted against the independent variables indicates that the linearity

assumption is met.
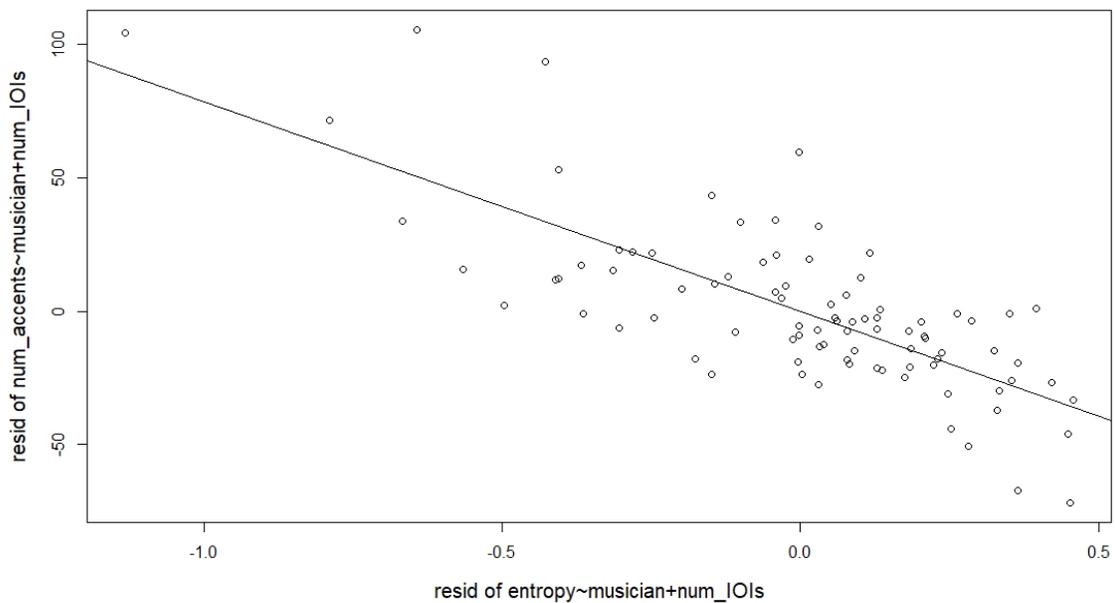


**Figure 21.** Added Variable Plot for number of accents

Recall that in the experimental portion of this project, I raised the issue of

multicollinearity, and found that the variables used there were not collinear. In the

present scenario, it is again desirable to verify the lack of collinearity in the data.

However, variance inflation factors (used in Chapter 4) only work if all of the variables in

a model have a single degree of freedom. In the present case, though, the variable

"musician" has *four* degrees of freedom (there are five musicians and one of them is

grouped together with the intercept). I can, however, use *generalized* variance inflation

factors. These rely on using "confidence ellipses or ellipsoids" rather than confidence

---

[14] Michael Lavine, in a personal communication, indicated that the test presented here is sufficient.

intervals for the variables that have more than a single degree of freedom. After calculating the GVIF's and transforming them using the formula $GVIF^{0.5v}$ where $v$ is the number of degrees of freedom, they can be compared across variables. They can be compared to the square root of the threshold value, usually 2.5, 5.0, or 10.0. In the present case, the transformed GVIF's are less than $\sqrt{2.5}$, so multicollinearity is not a problem here.
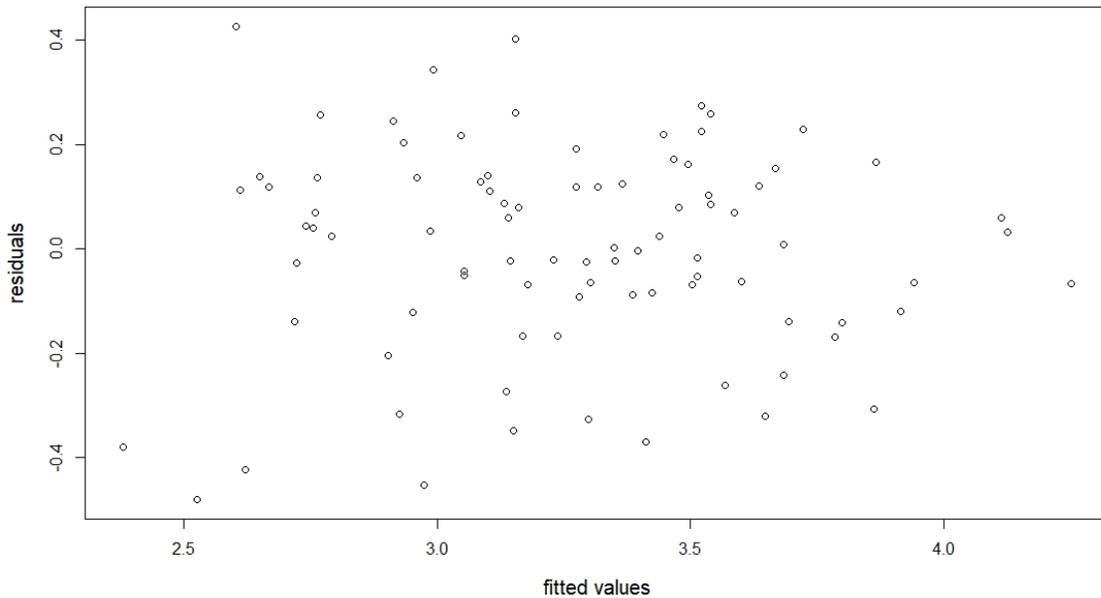


**Figure 22.** Residuals vs. Fitted Values Plot for two-covariate model

.

The final graphical technique, residuals vs. fitted values plots, tests whether or not a model is a good fit to the data, and whether or not more variables should be added. A complete, well-fitting model is indicated by a random scattering of points, showing that there is no systematic error in the data. The residuals vs. fitted values plot for the model *entropy ~ musician + num_IOIs + num_accents* is shown in Figure 22; it reveals that the model is a good fit to the data, and that adding more variables might result in overfitting.

The adjusted $R^2$ metric for this model was high: 0.7896. This indicates that most of the variation in entropy is due to the independent variables, rather than to noise.

In a situation such as this, when comparisons are made between multiple pairwise combinations, $p$ values must be adjusted to account for the multiple comparisons. That is because if each individual variable has a 0.95 confidence level, say, then for $n$ simultaneous pairs of variables the confidence level will be only $0.95^n$, necessarily less than 0.95. Increasing the individual $p$ values corrects this problem. The simplest and most reliable way to do this is to use the *Bonferroni correction*, which multiplies each pairwise $p$ value by the number of pairwise combinations. So in the present scenario, with five musicians, there are $5*4/2 = 10$ pairwise combinations, and we must multiply each unadjusted $p$ value by 10.

Here I digress for a moment to describe an anomaly in the data. In his solo on *I Never Knew*, Armstrong uncharacteristically plays a couple of clams, in measures 9 and 10, on beats four and one respectively. To examine the effects of these notes on entropy, I treated them four ways: neither as an accent; the first of the two as an accent; the second of the two as an accent; and both as accents. The net effect on entropy was less than five percent, the effect on num_IOI's was at most about eight percent (one out of twelve or thirteen), and the effect on num_accents was less than six percent (one or two in the range 33–35). None of these differences manifested as appreciable changes in $p$ values for the models discussed here, but to be as accurate as possible, $p$ values – as well as $R^2$values - were averaged across each of the four possible treatments of the clams[15][16].

---

[15] Note that the quantity in question is not perceptual, but intentional: what did Louis Armstrong *intend* to play? We can never know, so we average over the four possibilities.
[16] The added variable plots displayed above had almost no perceptible dependence on disposition of the clams.

One final topic pertaining to the analysis of this data must be addressed, as in the experimental case: interactions. As in the experimental case, two methods of checking for interactions are used: performing an analysis of variance on pairs of models having no interaction terms or having one or more interaction terms, and examining individual $p$-values for models having one or more interaction terms. The former method yields no interaction terms, while the second method yields an interaction between one value of the factor "musician" (specifically, the value "Parker") and number of IOIs. The difference in adjusted $R^2$ between the model with no interactions and the model with the significant interaction term is very small, however (<0.5%), so it is probably safe to ignore interactions in these calculations. (This is true for all dispositions of the clams).

## 5.4 Results

Table 9 shows the pairwise $p$ values resulting from applying the estimated marginal means technique; in each pair, the lower value is listed first. I find that while Armstrong, Christian, Hawkins, and Parker are all less entropic than Young, the difference between Hawkins and Young is not statistically significant since the $p$ value is greater than 0.05. In particular, the $p$ values for Armstrong/Young, Christian/Young, Hawkins/Young, and Parker/Young are 0.0224, 0.0239, 0.2404, and 0.0011. Pairs not involving Young all have $p$ values of 1.0, which occurs whenever multiplying a raw $p$ value by the Bonferroni correction results in a number greater than or equal to one.

| Musician Pair | *p Value* |
|---|---|
| Armstrong/Young | 0.0224 |
| Christian/Young | 0.0239 |
| Hawkins/Young | 0.2404 |
| Parker/Young | 0.0011 |

**Table 9.** Pairwise comparisons using the EMM procedure; comparisons not listed have *p* values of 1.0

Care must be taken in evaluating these results to take researcher error into account, since in a manually transcribed corpus of this size (6,299 accented notes; 20,000–30,000 notes total) there are bound to be errors. This is discussed in detail in Appendix C; the upshot is that random error probably does not qualitatively affect the above-reported *p* values, but there is a possibility that the Christian/Young comparisons should be reported as only *marginally* significant ($p \approx 0.05$–$0.06$), or even possibly insignificant. The pairwise comparisons between Armstrong and Young and Parker and Young, however, are secure.

I can take the analysis one step further by identifying outliers in the data. This is done by fitting the data with a multiple regression model and calculating residuals; points corresponding to residuals that are greater in absolute value than twice the standard deviation of the residuals are considered to be outliers. Using this method, five outliers were identified, including, notably, two early Charlie Parker excerpts, "Moten Swing" and "Honeysuckle Rose." "Honeysuckle Rose" will be discussed in greater detail later.

In light of the above discussion regarding Armstrong's solo on "I Never Knew," the identification of outliers was carried out for all four dispositions of the accented/non–accented clams. Results were consistent for all four treatments of the problematic notes.

Pairwise comparisons using the modified data are shown in Table 10. Removing outliers did not qualitatively change the results, but rather intensified the contrasts between Armstrong and Young, Christian and Young, and Parker and Young. *P* values were 0.0004, 0.0011, and 0.0006, respectively. The difference between Hawkins and Young was still not significant ($p = 0.2632$). All other $p$ values were 1.0 with the exception of Armstrong/Hawkins ($p = 0.4990$) and Christian/Hawkins (0.6673). The adjusted $R^2$ value for the data with outliers deleted was 0.818, even higher than that with outliers included. (Note that neither method of checking for interactions indicated that it was important to include interaction terms.)

| Musician Pair | *p Value* |
|---|---|
| Armstrong/Young | 0.00035 |
| Christian/Young | 0.0011 |
| Hawkins/Young | 0.2632 |
| Parker/Young | 0.0006 |
| Armstrong/Hawkins | 0.4990 |
| Christian/Hawkins | 0.6673 |

**Table 10.** Pairwise comparisons using the EMM procedure, data with outliers deleted. Pairs not shown have *p* values of 1.0

## 5.4 Debunking Hypotheses About Chronological Trends

Next I explore a way of parsing the data on Armstrong into early and late periods suggested by Barry Kernfeld (1996). Kernfeld posits that Armstrong's work from *before* 1936 is qualitatively different from his work from *after* 1936; he states that after 1936, Armstrong basically pandered to the masses, while his earlier work was more complex. I tested this hypothesis in terms of entropy using the EMM procedure, and found that there was no significant difference between the two periods ($p = 0.221$). Regarding Kernfield's assertion that Armstrong's later work pandered to the masses, one must only listen to his popular recordings of "What a Wonderful World" (1967), one of which (the one analyzed here) has an entropy value of 3.877, to see that Armstrong could gain the widespread adulation of fans without sacrificing musical complexity.

Again using the EMM technique, I explore the commonly held belief that Lester Young's playing was altered by his stint in the Army, and find that entropy does not reveal a difference between his pre- and post-Army solos ($p = 0.229$).

## 5.5 Hawkins vs. Parker

The fact that the pairwise $p$ value for Hawkins and Parker is not even close to significance (it is one, actually) merits some attention; this result is counterintuitive, given that Parker is usually considered to be far more rhythmically complex – and therefore more entropic – than Hawkins.

For example, according to Scott DeVeaux in The Birth of Bebop, Hawkins was known for his rhythmic predictability:

*His notorious tendency toward rhythmic uniformity – a steady stream of eighth notes …*
*later caricatured by one writer as a "machine-gun style" – was only exacerbated by his*

*ongoing project of crowding as much of the underlying harmony as possible into his*

*improvised line.*

<div align="right">DeVeaux 1997, 85</div>

Martin Williams confirms this observation:

*Rhythmically, [Hawkins] continued to live in the early 'thirties – but, again, with more*

*regular accents than many players of that period.*

<div align="right">Williams 1993, 77</div>

Parker, on the other hand, was known for his rhythmic adventurousness:

*... the pattern of accents in a Charlie Parker line is in a constant state of flux – falling*

*sometimes on the strong beats of the measure, but also (and quite unpredictably) on*

*"weak beats" (beats 2 and 4) or on the weak half of the eighth note pair.*

<div align="right">DeVeaux 1997, 264</div>

If entropy does not reflect the predictability of Hawkins and unpredictability of

Parker we are led to expect from conventional wisdom, what accounts for these

expectations? One possibility is that Hawkins uses more dynamical accents than Parker,

perhaps creating an expectation of accents. This theory is supported by the fact that the

average dynamical accent density (number of dynamical accents divided by excerpt

length) for a limited corpus of thirteen Hawkins excerpts is 1.74 while for a corpus of

fifteen Parker excerpts it is 1.36[17]. A two-sample $t$-test yields a $p$ value of 0.0046; so the

difference is statistically significant.

---

[17] The distribution of dynamical accents for a corpus of sixteen Parker excerpts was found not to be normal, a requirement for using the two-sample $t$ test used here. Deleting a single excerpt ("Hootie Blues") resulted in a normal distribution.

I can go one step further, and obtain results based not only on *dynamical* accents but on *contour* accents as well[18]. Contour accents occur when the melody changes direction; only changes in direction prepared by two intervals moving in the same direction were counted. This is a reasonable thing to do, given that, to my ears, Parker's lines owe much of their interest to changes in direction. For example, the lick shown in Figure 23 changes direction eight times in the space of just two bars. This lick appears in "Au Privave," "Billie's Bounce," and "Now's The Time." (To be fair, this lick also appears once in a Hawkins solo, but not frequently as in the Parker corpus). Using the same samples as those used for the dynamical accent calculations[19], I find that the average contour accent density for Hawkins is 1.03, while for Parker it is 1.34. Again using a two-sample $t$-test, I find that $p = 0.00078$. Thus, there is a significant difference in contour accent density between Parker and Hawkins, perhaps creating the expectation of unpredictability, unlike the expectation created by greater density of dynamical accents.



**Figure 23.** Charlie Parker lick changing direction eight times in two bars

Recall from Chapter 4 that entropy, to the extent that we can assume it affects the perception of rhythmic complexity for rhythms that are not eighth-note-based, is not the only factor affecting the perception of rhythmic complexity. Periodicity (or lack thereof),

---

[18]For more on contour theory, see Deutsch (1972), Dowling (1978), Edworthy(1985), Friedmann (1985), Polansky (1996), Quinn (1999), and Schmuckler (1999).
[19]Including or excluding "Hootie's Blues" did not affect the normality of the contour accent sample, so it was included in the two-sample $t$ test.

syncopation, and number of notes (for an experiment involving approximately equal excerpt lengths), for eighth-note-based rhythms, are all factors of interest. Recall that in Chapter 4 I pointed out that it cannot be decided *a priori* whether listeners perceive note *number* or note *density.* For the present argument I will assume that it is *density* that is perceived; bear in mind that this may be a faulty argument.

Note that in designing the estimated marginal means model for entropy, accent number was used as a covariate rather than accent density because I desired a variable that reflected length of excerpt and because the added variable plot clearly indicated that it made sense to add number of accents to the independent variable "musician" and the covariate "number of distinct IOIs". In investigating number of notes or note density as factors of interest leading to perceived rhythmic complexity, however, density is preferred because this is an intuitive way of normalizing complexity, while normalizing entropy by number of notes absolutely is *not*.

To take this hypothetical situation one step further, let us accept the suppositions that entropy depends on number *count* (among other things), and that perceived rhythmic complexity depends on entropy and number *density* (among other things). If complexity depends on entropy, and entropy depends on number count, then we might naively expect complexity to depend on number count, too. To resolve this conundrum, I use the variance inflation factor (V.I.F.) introduced in Chapter 4. Recall that in the model *complexity ~ entropy + number/density of notes + order effects*, the V.I.F. tests for multicollinearity in the predictor variables, and finds none. Entropy and number/density of notes (which are approximately the same in the experiment setup used here) tell us different things about the data. We need not concern ourselves with the propagation of

the choice between number count and number density from the entropy calculation to the perceived complexity calculation. In other words, we don't need to know what the exact relationship is between number (density or count) and entropy in the perceived complexity calculation. All we need to know is revealed by the estimated marginal means procedure for computational results and the linear mixed effects regression procedure for experimental results.

It should be noted, however, that the absence of triplet- or sixteenth-note-based rhythms or ornamental straight-eighth-notes in the experimental excerpts may impact the application of note density to any theorizing about how note density affects the perception of rhythmic complexity. For example, a subsequent experiment might indicate that quarter-note triplets impact one's perception of rhythmic complexity more, or at least differently, than eighth-notes do.

## 5.6 Conclusion

In this chapter I tested the hypothesis that computed entropy depends on musician in a corpus of 88 solos by Armstrong, Hawkins, Young, Christian, and Parker using the estimated marginal means technique with number of IOIs and number of accents as covariates. Using this technique obviated the need for normalization or minimum sample size. Furthermore, the added variable plot technique strongly supported the inclusion of these covariates.

Results showed that solos by Young were significantly more entropic than solos by Armstrong or Parker, and *probably* more entropic than solos by Christian (the

presence of researcher error made it impossible to ascertain the results regarding Christian for sure).

Two hypotheses regarding chronological trends were shown not to be supported by entropy calculations.

Finally, an alternative explanation for the fact that Parker was not found to be more entropic than Hawkins was sought by comparing dynamical accent density and contour accent density between the two soloists: Hawkins was found to have higher dynamic accent density while Parker was found to have greater contour accent density. This at least partially explains this result.

**CHAPTER 6**

**CLOSE READINGS**

**6.1 Introduction**

While Chapter 5 dealt with data in the aggregate, i.e. solos grouped by musician, in this chapter I examine five pairs of excerpts and one individual excerpt in order to get a better feel for how entropy works beyond the context of the estimated marginal means procedure. In order to do this, I employ several metrics: 1) the ratio of accented notes that fall on downbeats (DBR) or on strong beats (beats one and three, SBR) to the number of measures; 2) the syncopation metric of Longuet-Higgins and Lee in its original form, divided here by the number of accents to yield a quotient (LHLQ); 3) accent density (number of accents divided by number of measures); 4) note density (notes per measure); and 5) the Kullback-Leibler distance (KLD) from the uniform distribution with as many IOIs as the excerpt in question. Tables 11a and 11b illustrate the values of the above-mentioned metrics, as well as their corresponding number of IOI's, number of accents, and entropy values.

Pairs were selected in order to make meaningful comparisons: the first four pairs are grouped according to similar or exactly the same values of the two covariates, and the fifth pair corresponds to the first 32 bars of two interpretations of "Body and Soul" played by Hawkins and Young. The unpaired excerpt, "Blues For Greasy" by Young, has the highest entropy in the corpus (4.186).

The first 32 bars of "Body and Soul" are isolated because Hawkins' version consists of two contiguous 32 bar choruses (with an ornamented rubato ending), while

Young's version consists of just 32 contiguous bars followed by piano and bass solos,

another 16 contiguous bars, an eight-bar piano solo, and an 8 bar ending.

| Musician | Excerpt Name | n_IOIs | n_acc | DBR | SBR | LHLQ |
|---|---|---|---|---|---|---|
| Parker | Honeysuckle Rose | 10 | 62 | 0.6875 | 1.4375 | 0.726 |
| Christian | Flying Home | 10 | 61 | 0.5 | 1 | 1.525 |
| Armstrong | Hello Dolly | 13 | 39 | 0.6465 | 1.0101 | 1.103 |
| Christian | Gone With What Wind | 13 | 39 | 0.4123 | 0.701 | 1.974 |
| Parker | Ornithology | 17 | 43 | 0.2813 | 0.5938 | 2.535 |
| Christian | Seven Come Eleven | 17 | 43 | 0.2796 | 0.5903 | 2.395 |
| Young | Tea For Two 1947 | 17 | 82 | 0.2813 | 0.6406 | 2.366 |
| Hawkins | Epistrophy | 18 | 82 | 0.4063 | 0.75 | 1.329 |
| Hawkins | Body and Soul 1st chorus | 16 | 103 | 0.4563 | 0.835 | n/a |
| Young | Body and Soul 1st chorus | 19 | 73 | 0.2027 | 0.3108 | n/a |
| Young | Blues For Greasy | 21 | 38 | 0.3611 | 0.5278 | 1.947 |

**Table 11a.** Number of IOIs, number of accents, downbeat ratio, strong beat ratio, and LHL quotient for four pairs grouped by covariates, one pair grouped by song, and one individual excerpt ("Blues For Greasy")

| Musician | Excerpt Name | note dens | accent dens | KLD | Entropy |
|---|---|---|---|---|---|
| Parker | Honeysuckle Rose | 653.13% | 193.75% | .778/.769/.774 | 2.199 |
| Christian | Flying Home | 475.00% | 190.63% | .369/.455/.412 | 2.789 |
| Armstrong | Hello Dolly | 351.72% | 156.00% | .622/.553/.588 | 2.803 |
| Christian | Gone With What Wind | 495.83% | 162.50% | .333/.315/.324 | 3.22 |
| Parker | Ornithology | 609.38% | 134.34% | .434/.356/.395 | 3.461 |
| Christian | Seven Come Eleven | 466.67% | 130.30% | .312/.287/.299 | 3.638 |
| Young | Tea For Two 1947 | 396.88% | 126.15% | .452/.424/.438 | 3.435 |
| Hawkins | Epistrophy | 498.44% | 126.15% | .350/.347/.348 | 3.665 |
| Hawkins | Body and Soul 1st chorus | 470.04% | 160.31% | .960/.924/.942 | 2.616 |
| Young | Body and Soul 1st chorus | 346.88% | 114.06% | .438/.396/.417 | 3.616 |
| Young | Blues For Greasy | 429.73% | 102.70% | .143/.135/.139 | 4.186 |

**Table 11b.** Note density, accent density, Kullback-Leibler Density, and entropy for four pairs grouped by covariates, one pair grouped by song, and one individual excerpt ("Blues For Greasy")

Several observations about this data will be made. First of all, since entropy and

syncopation are conceptually different quantities, there is sometimes a correlation

between the two, and sometimes there is not. The above-defined quantities DBR and SBR, on the other hand are, with few exceptions, inversely correlated with syncopation ($r$ = –0.92 and –0.89 respectively), since syncopation by definition depends on the presence of off-beat notes. KLD, with few exceptions (compare Hawkins "Body and Soul" to Parker "Honeysuckle Rose"), is inversely correlated with entropy. The point corresponding to Hawkins' rendition of "Body and Soul" is clearly an outlier and strongly affects the correlation value: correlation scores are $r$ = –0.74 with the outlier present, –0.89 with the outlier deleted. This makes sense since a low KLD indicates closeness to a uniform distribution and a high entropy. Within pairs, these relationships are even stronger, with only a single pair ("Ornithology" and "Seven Come Eleven") bucking the trend between DBR and LHLQ. Intriguingly, the density of accented notes seems to be quite similar within each pair. This result merits further investigation.

Hawkins and Young are compared in the context of their respective solos on "Body and Soul." And finally, the high entropy of Young's solo on "Blues For Greasy" is qualitatively explicated.


## 6. 2 Pairs Determined by Covariates

Several patterns can be seen here. To begin with, note that the excerpts within the first four pairs (Parker's "Honeysuckle Rose"/Christian's "Flying Home," Armstrong "Hello Dolly"/Christian's "Gone With 'What' Wind," Parker's "Ornithology"/Christian "Seven Come Eleven," Young's "Tea For Two" (1947)/Hawkins's "Epistrophy") are ordered from low entropy to high. Note, too, that the entropies between the first three

pairs are ordered from low to high, while the entropies within the fourth pair overlap those within the third pair.

Within the first two pairs, the first excerpt has a higher ratio of accents falling on down beats or strong beats, lower normalized syncopation scores, a higher KLD, and a lower entropy. These relationships can be explicated as follows.

As expected, the percentage of accents falling on down beats or strong beats is negatively correlated with the LHL syncopation metric, since syncopation is by definition greater when the accents fall further from downbeats or strong beats. It also makes sense that the percentage of accents falling on downbeats or strong beats would be *positively* correlated with the Kullback-Leibler distance from uniformity. This is because most excerpts emphasize three IOIs in particular, to a greater or lesser degree: two, four, or eight eighth notes, corresponding to quarter notes, half notes, or whole notes. Sometimes they fall between accented off-beats, for example, between the and of one and the and of three, but more often they fall between quarter note beats one, two, three, or four. The most common reason for high KLD, then, is a predominance of one or more of these IOIs, notably including beats one or three; thus, the positive correlation between DBR, SBR, and KLD.

To take an extreme example, examine the transcription of Parker's early (1940) recording of "Honeysuckle Rose" (Figure 24). Note the overwhelming frequency of accented strong beats (one and three). This tendency to align accents with strong beats is clearly seen in the probably distribution, as shown in Figure 25: there is a sharp peak at an IOI of four eighth notes, which reflects the predominance of beats one and three in the solo.

Since this performance takes place at an extremely fast tempo (about 280 bpm), one might wonder whether or not the predominance of accents on beats one and three is due to the soloist's inability to articulate accents on weak beats at this tempo. However,

## Early Parker Honeysuckle Rose



**Figure 24.** Transcription of Charlie Parker's solo on "Honeysuckle Rose"

examining Parker's mature recording of "KoKo" (1945) at a similar (slightly faster) tempo shows this not to be the case. Utilizing the same metrics as those used to evaluate "Honeysuckle Rose," DBR and SBR, I find values of 0.4141 and 0.8586 (for "Honeysuckle Rose" these values are 0.6875 and 1.4375) respectively. So while tempo may play a role in high DBR and SBR ratings for "Honeysuckle Rose," it does not always interfere with emphasizing off-beats. As demonstrated by "KoKo," he was in fact able to emphasize off-beats, even at fast tempos – at least later in his career.



**Parker Honeysuckle Rose**
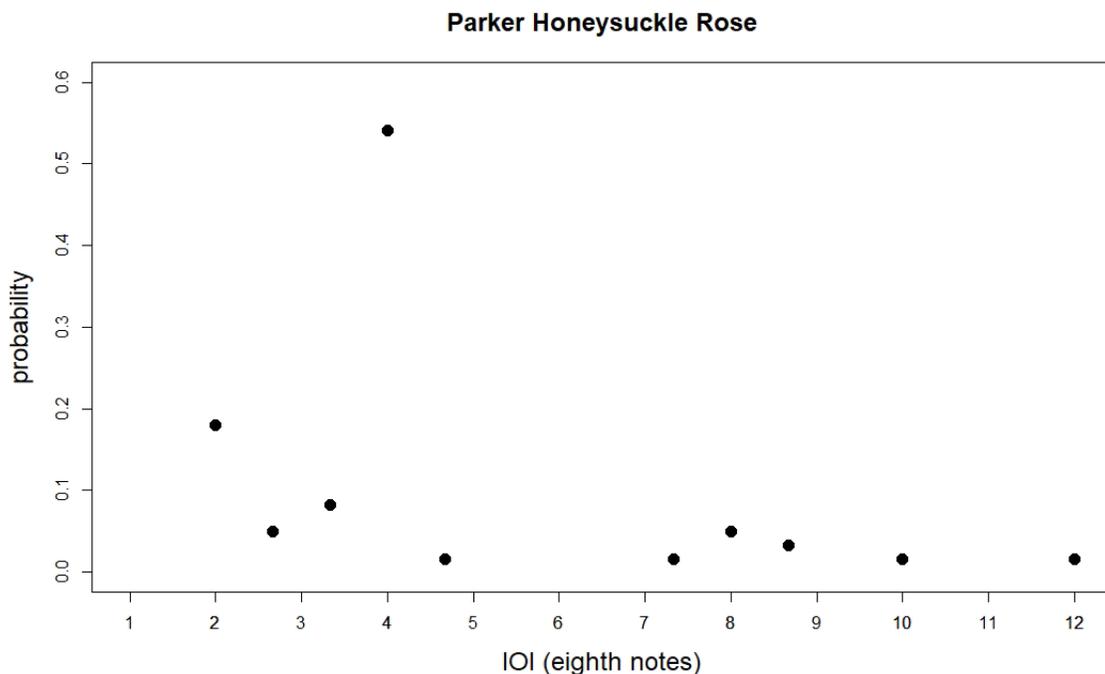
**Figure 25.** Probability distribution for Charlie Parker's "Honeysuckle Rose"

By contrast, Christian's solo on "Flying Home" (transcription Figure 26, distribution Figure 27) does not show a strong preference for IOIs of 2, 4, or 8 so there are no peaks in the distribution of the same order of magnitude as in Parker's solo on "Honeysuckle Rose." Also, the DBR and SBR are lower (this does not necessarily follow from the above observation); the fact that the DBR and SBR are lower means that

the LHL quotients are higher, the KLD is lower, and the entropy is higher. This pair

evokes the contrast between the uniform distribution and the one-sharp-peak distribution

examples in Chapter 2. So we see that the Christian excerpt is rhythmically freer than the

early Parker excerpt, an interesting result because Parker's solos are frequently

considered to be the apogee of rhythmic unpredictability.

## Christian Flying Home



**Figure 26.** Transcription of Charlie Christian's solo on "Flying Home"
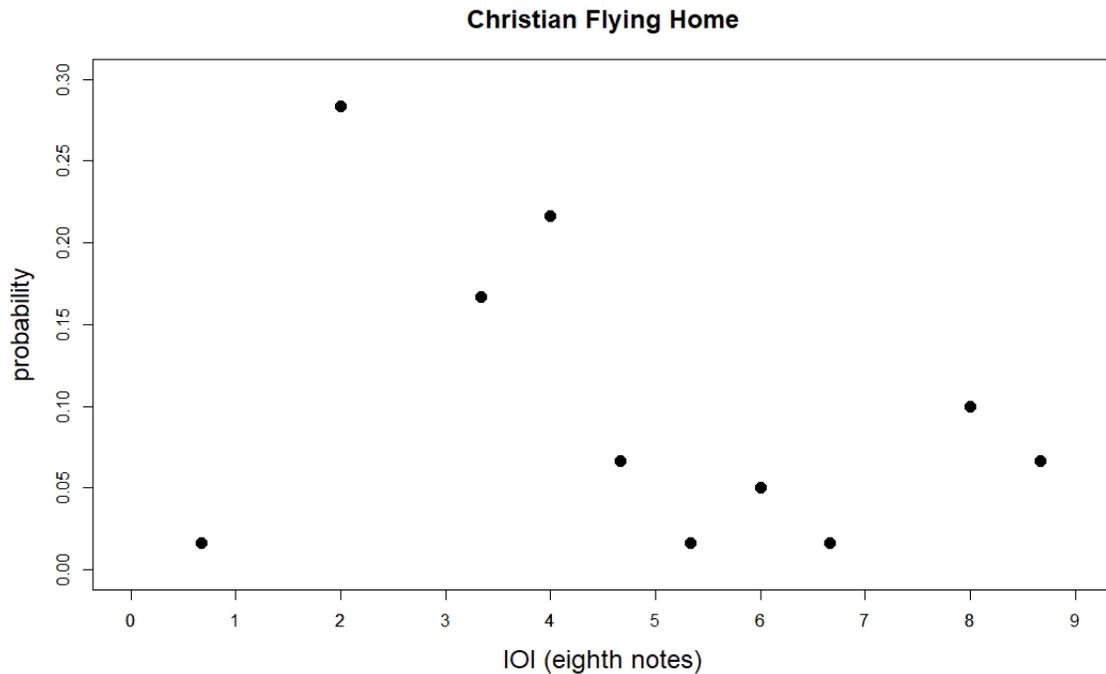
**Figure 27.** Probability distribution for Charlie Christian's solo on "Flying Home"

Comparing Armstrong's solo on "Hello, Dolly" (transcription Figure 28,

distribution Figure 29) to Christian's solo on "Gone With 'What' Wind" (transcription

Figure 30, distribution Figure 31) evokes the contrast between the second and fourth

sample distributions in Chapter 2: one that exhibits a single peak and another that exhibits

two peaks of magnitude roughly equal to that of the single peak. In particular, the

distribution for "Hello, Dolly" has two peaks with magnitudes roughly equal to 0.3, while

the distribution for "Gone With 'What' Wind" has a *single* peak, also equal to 0.3. As

explained in Chapter 2, the greater amount of total probability in the two peaks in the

former results in a higher KLD and lower entropy than the amount of probability

contained in the single peak in the latter. Thus, the greater amount of rhythmic variability

in the second excerpt results in higher entropy. It is noteworthy that the Armstrong

excerpt is probably what Barry Kernfeld had in mind when he stated that later Armstrong

was sparser; thus it is not surprising that Armstrong's solo on "Hello, Dolly" would be less entropic than Christian's "Gone With 'What' Wind."

## Armstrong Hello Dolly



**Figure 28.** Transcription of Armstrong's solo on "Hello, Dolly"

The third pair of excerpts – Parker's solo on "Ornithology" (transcription Figure 32, distribution Figure 33) and Christian's solo on "Seven Come Eleven" (transcription Figure 34, distribution Figure 35) – displays a contrast between a distribution with a single sharp peak of magnitude 0.3 ("Ornithology"), and a distribution with a much less pronounced peak of magnitude 0.2 ("Seven Come Eleven"). The Parker excerpt therefore has a higher KLD and lower entropy than the Christian excerpt. Note, however, that,

unlike the first two pairs, the strong-beat ratio for the second excerpt is greater than that

**Armstrong Hello Dolly**
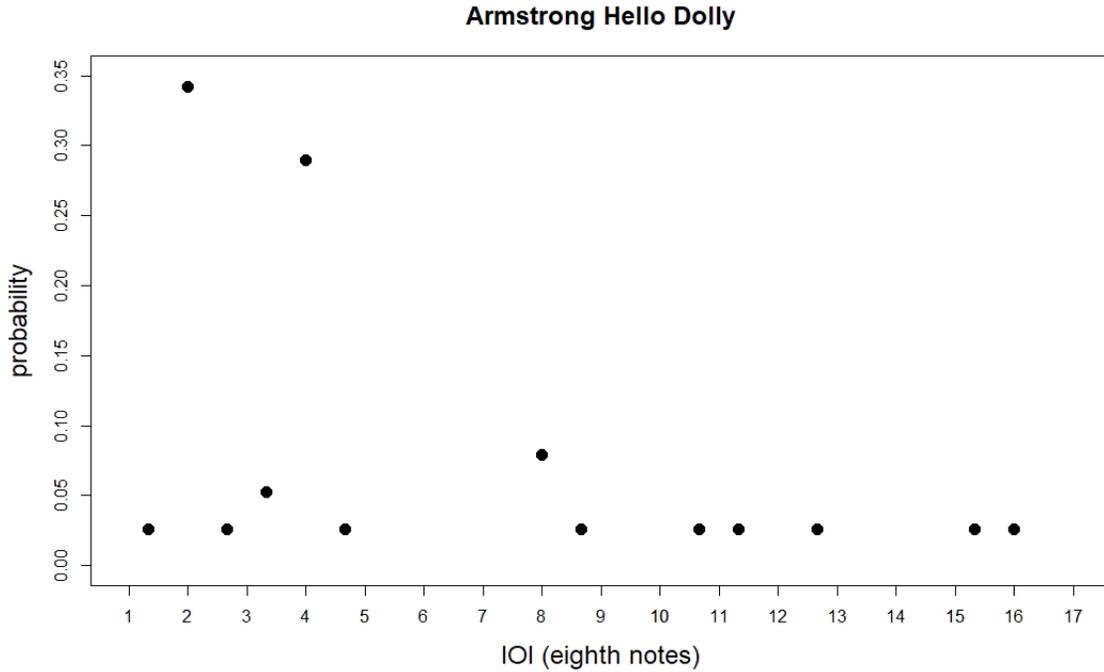


**Figure 29.** Probability distribution for Armstrong's solo on "Hello Dolly"

for the first excerpt. This correlates with the fact that the LHLQ score for "Seven Come Eleven" is lower than that for "Ornithology." None of this contradicts the fact that the entropy is higher for the second excerpt than for the first, since entropy is independent of meter. Note, too, that KLD is lower for the second excerpt than for the first, in keeping with the fact that the relationship between KLD and entropy is consistent: higher KLD means lower entropy and vice-versa.

The fourth pair – Young's 1947 solo on "Tea For Two" (transcription Figure 36, distribution Figure 37) and Hawkins' solo on "Epistrophy" (transcription Figure 38, distribution Figure 39) exhibits a similar situation as in the third pair. In particular, both the DBR and SBR are higher for "Epistrophy" than for "Tea For Two," and LHLQ scores

are lower for "Epistrophy" than for "Tea For Two." These observations are consistent with one another and independent of the fact that the entropy of the second excerpt is greater than that of the first excerpt, since entropy is independent of meter. A closer inspection reveals some interesting patterns.

## Christian Gone With "What" Wind



**Figure 30.** Transcription of Christian's solo on "Gone With 'What' Wind"

First of all, the distribution for "Tea For Two" reveals a single sharp peak for an IOI of four eighth-notes. This would tend to decrease the entropy. Why, then, should the syncopation for "Tea For Two" be greater than that for Hawkins's "Epistrophy"? An

inspection of the "Tea For Two" transcription reveals that IOIs of four eighth-notes sometimes begin on weak beats (m. 4–5, m. 15, m. 20, m.30–31), yielding two syncopations for each IOI beginning on a weak beat. IOIs of four eighth-notes beginning on beats 1 or 3 – and ONLY those beginning on beats 1 or 3 – do not yield any syncopations. Thus, syncopation can be relatively high even when entropy is relatively low.

Note that Hawkins begins his solo by quoting the melody. By virtue of the melody's structure, there are four instances of IOI six (m. 3–4, m. 5–6, m. 7–8, m. 11–12); IOI six appears frequently thereafter, as well. Not surprisingly, IOI four occurs frequently as well, as it does in many other excerpts (see above). Nonetheless, the two peaks in the distribution are not large enough to lower the entropy below the entropy of the preceding excerpt (Young's 1947 "Tea For Two").
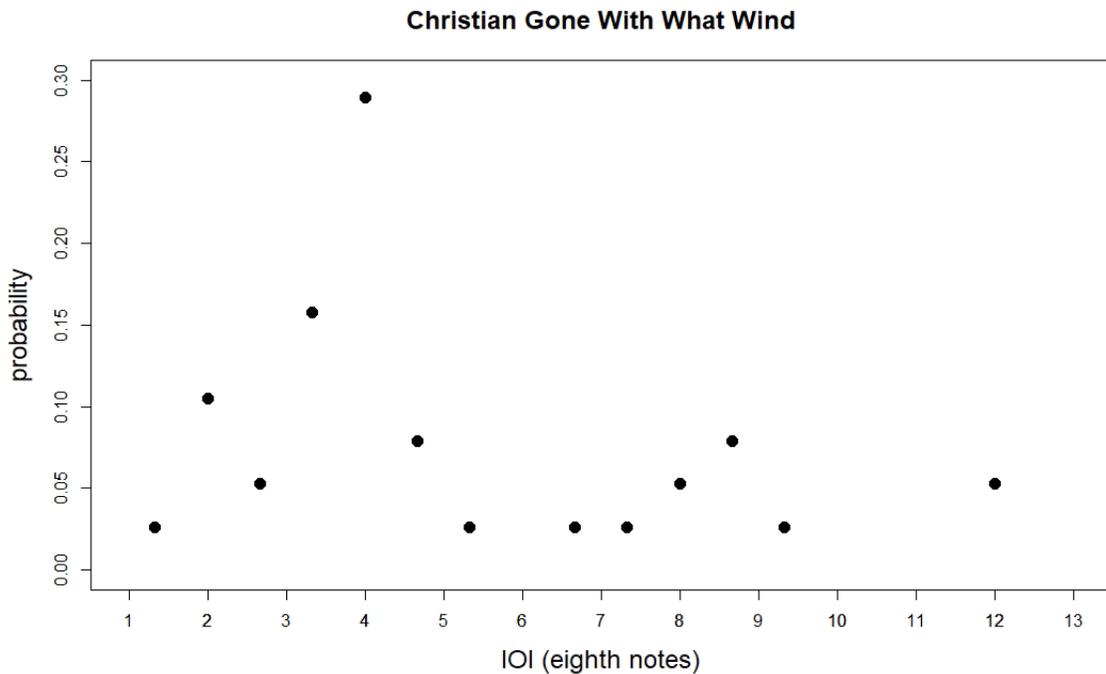


**Figure 31.** Probability distribution for Charlie Christian's solo on "Gone With 'What' Wind

**Figure 32.** Charlie Parker's solo on Ornithology
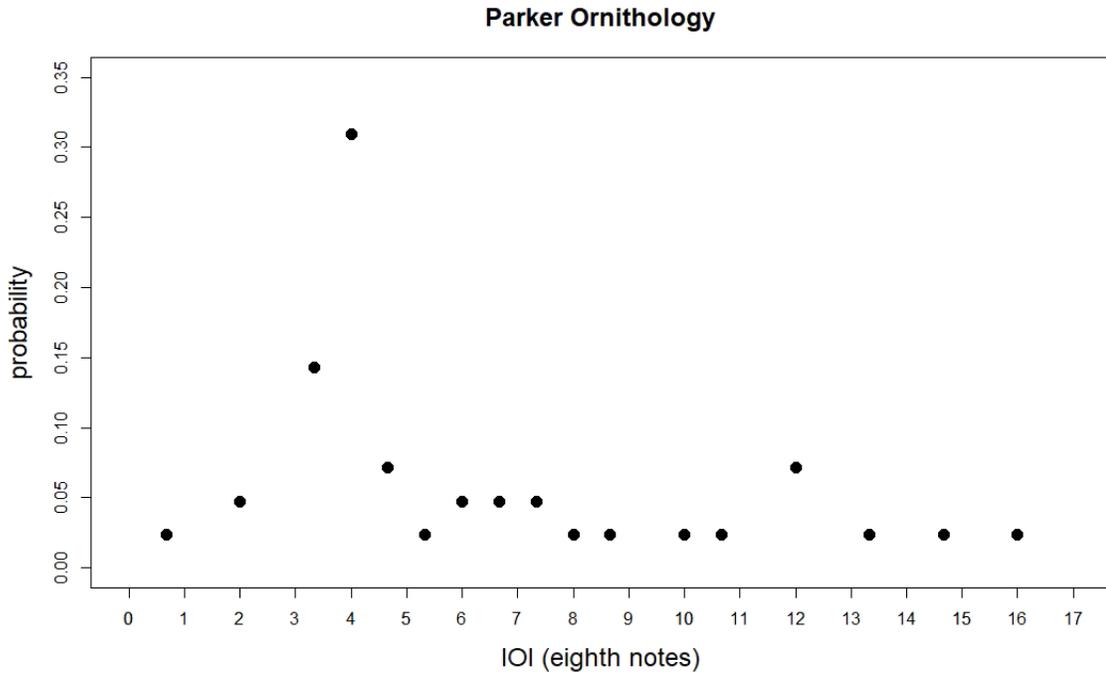
**Figure 33.** Probability distribution for Charlie Parkers's solo on "Ornithology"



**Figure 34a.** Charlie Christian's solo on "Seven Come Eleven" (1 of 2)

**Figure 34b.** Charlie Christian's solo on "Seven Come Eleven" (2 of 2)
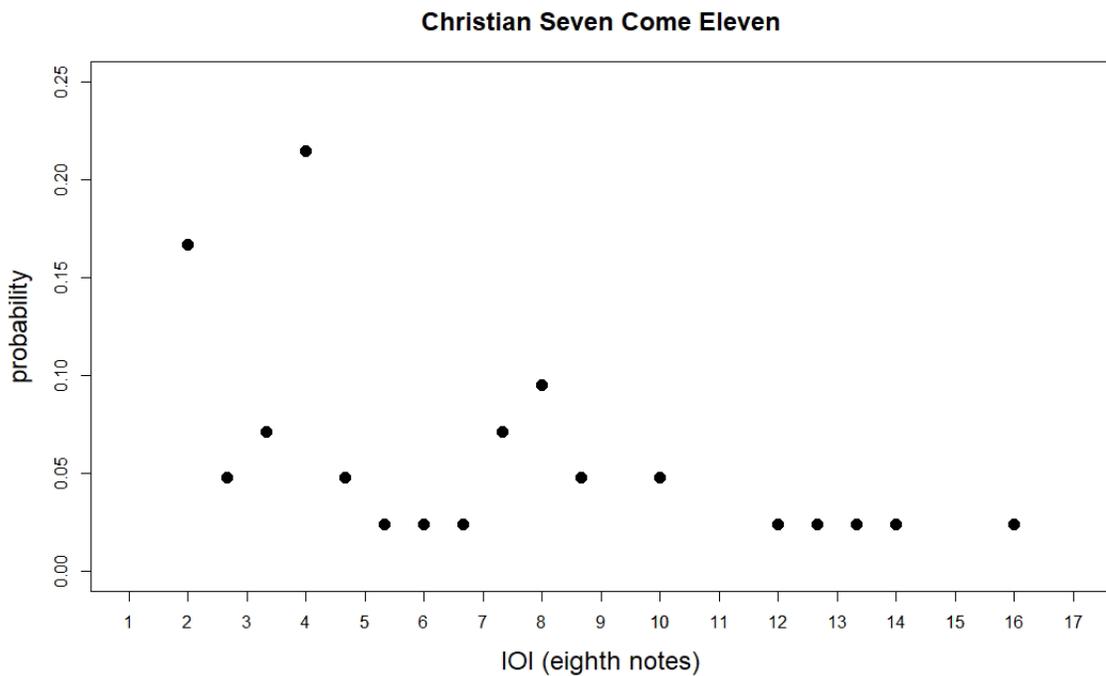


**Figure 35.** Probability distribution for Christian's solo on "Seven Come Eleven"

# Young
## Tea For Two 1947



**Figure 36a.** Transcription of Young's solo on "Tea For Two" (1947) (1 of 2)

**Figure 36b.** Transcription of Young's solo on "Tea For Two" (1947) (2 of 2)

**Young Tea For Two 1947**
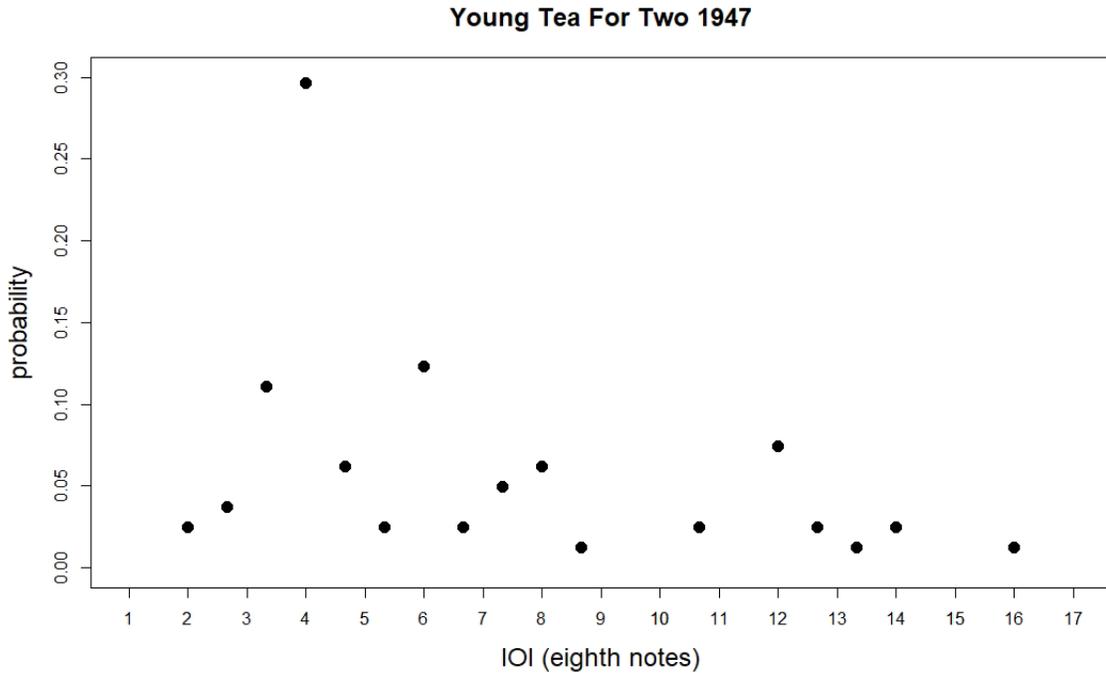


**Figure 37.** Probability distribution for Young's solo on "Tea For Two" (1947)



**Figure 38a.** Transcription of Hawkins's solo on "Epistrophy" (1 of 3)

**Figure 38b.** Transcription of Hawkins's solo on "Epistropy" (2 of 3)

110

**Figure 38c.** Transcription of Hawkins's solo on "Epistrophy" (3 of 3)
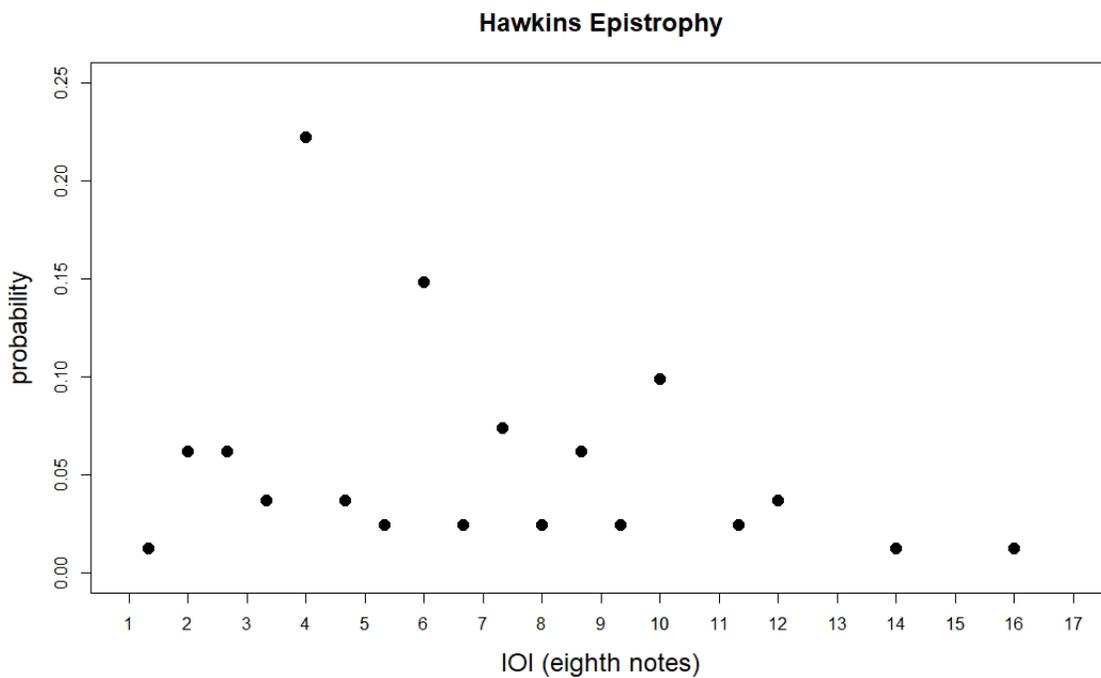


**Figure 39.** Probability distribution for Hawkins's solo on "Epistrophy"

**6.3 Hawkins vs. Young**

Here I investigate two commonly held beliefs about Hawkins and Young by analyzing their interpretations of "Body and Soul." First, I look into the commonly held notion that Young is more laconic in his playing than Hawkins. Second, I explore the idea that Hawkins is a "vertical" player and Young is a "horizontal" player.

Gunther Schuller, discussing the style of trumpeter Oran "Hot Lips" Page in the 1930's, posits that

*Page's best solos reflect a minimum of activity with a maximum of expression, a lesson Lester Young was to extend several years later.*

Schuller 1968, 316

Coleman Hawkins, on the other hand, is usually considered to be a verbose, rhythmically uniform player, the "hot" to Lester Young's "cool." According to Martin Williams,

*[Hawkins's] chorus on "Dinah" … frankly sets up the expectation of more or less regular heavy/light/heavy/light accents and varies them just briefly enough, often enough, and obviously enough to relieve any encroaching monotony.*

Williams 1993, 75

In fact, Young is often considered more laconic than most players, and Hawkins is often considered more verbose than most players. Here I explore the notion that Lester Young's style involved fewer notes than Hawkins using the admittedly small sample size of two renditions of *Body and Soul*, one the famous 1939 version by Hawkins, and the other a 1945 version by Young at the same tempo.

This is, on the whole, borne out by these transcriptions: the first 32 bars of

Young's solo contains 222 notes, while the first 32 bars of Hawkins' contains 292. This

is a difference of about 27%. The next sixteen bars of Young's solo (after one chorus of

piano and bass solos) contains 135 notes, while the next sixteen (contiguous) bars of

Hawkins' solo contains 148 notes. Finally, the last eight bars of Young's solo (after an

eight-measure piano solo) contain 66 notes, while the last eight bars of Hawkins' solo

contain 59 notes. Summing these note counts gives us, for Young, 423, and for Hawkins,

499, a difference of about 16%. This seems to confirm the notion that Hawkins is more

verbose than Young, though this is not a conclusive result.

Next I investigate the commonly held notion that Hawkins is a more vertical or

harmonic player than Young, who is usually considered to be a more horizontal or

melodic player. Quoting Jonathan Mayhew, from a blog sponsored by the Stanford

Humanities Center,

*Some writers refer to horizontal (melodic) improvisation vs. vertical (harmonic)*

*improvisation.  The melodic tradition is that of Lester Young, Sonny Rollins, and Ornette*

*Coleman.  The harmonic tradition reaches its apogee in Coleman Hawkins and John*

*Coltrane.*

<div align="right">Mayhew 2022</div>

Or quoting Williams again:

*Young often overrode harmony in the interests of melody and his original rhythmic ideas.*

<div align="right">Williams 93, 77</div>

Naively, we can scrutinize this by counting arpeggios. For present purposes, I will define

"arpeggio" to mean three or more adjacent chord members related by skip or by step (in

the case of seventh chords), a triad with a note filling in between root and third, or broken arpeggios such as 1–5, 1–3, 3–5. I find that, in the first thirty-two bars, there are 30 arpeggios in Young's solo and 26 in Hawkins' solo. In the remaining (non-contiguous) 24 bars of Young's solo there are 4 arpeggios, while in the A sections of Hawkins' second chorus, there are 10. So altogether, there are 34 arpeggios in Young's version of Body and Soul, and 36 arpeggios in Hawkins' version. This would not seem to be a clear difference between Hawkins and Young in terms of vertical vs. horizontal playing. However, counting arpeggios is not the only way of quantifying vertical vs. horizontal playing.

In order to understand the differences in Hawkins' and Young's playing in terms of the vertical/horizontal paradigm, I turn to a Schenkerian concept: compound melody (see e.g. Forte and Gilbert 1982, p. 67 ff.). Roughly speaking, compound melody occurs when a single voice implies two or more individual melodies separated by register. Hawkins uses compound melody frequently. For example, compound melodies can be found in m. 2, m. 10–11, m. 26–7, m. 28–9, m. 35–6, m. 49–50, m. 55–6, and m. 57–9. One example, from m. 26–7, is shown in Figure 40. This is typical of Hawkins' playing; in particular, what these examples have in common is that Hawkins leaps from one register to another, specifically from a higher register to a lower register, often with the higher note receiving a short duration, then leaps up again and down again to complete stepwise motion in both registers. Young does not do this in his solo on "Body and Soul," though he does play a version of it in m. 15–16 (Figure 41).

**Figure 40.** Example of compound melody from Hawkins, "Body and Soul" (m. 15–6)



**Figure 41.** Example of compound melody from Young, "Body and Soul" (m. 26–7)

It should come as no surprise that Hawkins employs compound melody so frequently. He played 'cello as a child, and in an interview in *The New York Daily News* in 1965 said that he spent at least two hours every day listening to Bach (Chilton 1990, 366); surely he must have been familiar with the Bach unaccompanied cello suites, which include prime examples of compound melody.

### 6.3 "Blues For Greasy"

One of the most striking features of Lester Young's solo on this jam session piece (transcription Figure 38) is the prominence of the accented notes. Perhaps this is correlated with the low accent density (102.70%, the lowest of the eleven excerpts considered here). This is in keeping with Schuller's observation that Young often created "maximum expression" with "minimum material." One way this economy of means affects the entropy is through the fact that large IOI numbers are possible, for example: 10,12, 14, 15, or 16. This in turn facilitates a more even probability distribution.

**Figure 42.** Transcription of Young's solo on "Blues For Greasy"

116

Furthermore, Young frequently exploits the built-in rhythmic complexity in uneven eighth-notes to facilitate a more even probability distribution. He does this by starting with an IOI containing an odd number of eighth-notes and starting a second IOI on the ending note of the first one, containing the same number of eighth-notes. This makes for two different IOI lengths, despite the fact that they both contain the same number of eighth-notes, since one will contain short/long alternations SL…S and the second will contain LS…L. Examples contain 1 eighth-note (m. 13), 3 eighth-notes (m. 14–5), 7 eighth-notes (m. 5–6), and 9 eighth-notes (m. 21–23). (He also repeats an IOI of 16, which does not exhibit this property because the number of eighth-notes it contains is even; m. 32–33).
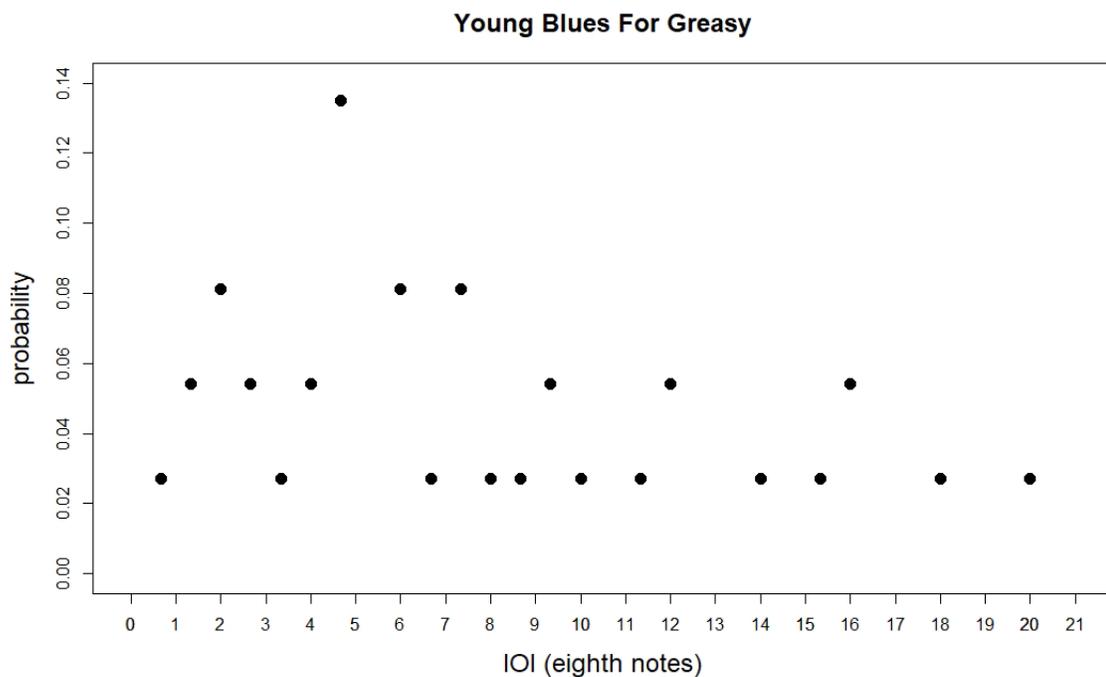


**Figure 43.** Probability distribution for Young's solo on "Blues For Greasy"

More than anything else, however, this excerpt has such a high entropy because there are simply no large peaks in the probability distribution (Figure 43). As mentioned

117

previously, most solos feature prominent peaks at 2, 4, or 8 eighth-notes, resulting in substantial Kullback-Leibler distance from the uniform distribution and a low entropy. This solo lacks such peaks – i.e. Young avoids accent pairs separated by 2, 4, or 8 eighth-notes - and thus closely approximates a uniform distribution, hence having a high entropy.

# CHAPTER 7

# ENTROPY AND MELODIC EMBELLISHMENT

A vital part of the jazz tradition is the embellishment of standard songs to become jazz. Examples can be found in the heads corresponding to the improvisations in the corpus used here, such as Young's renditions of "All of Me" and "Tea For Two." In other cases, such as Hawkins's famous 1939 rendition of "Body and Soul" (see chapter 6), or Armstrong's renditions of "Stardust" (1931) and "What a Wonderful World" (1967), the embellishments are so extreme as to render the heads part of the solos themselves.

Melodic embellishment is one of the four types of improvisation enumerated by Henry Martin in his study of thematic improvisation in Charlie Parker's playing (1996, 34), based on classifications devised by Kernfeld and Hodier – namely, paraphrase improvisation. The other three types are chorus phrase improvisation (based on the form and harmonic structure of the head only), motivic improvisation (based on motives from the head), and formulaic improvisation (based on formulas used by the soloist throughout his or her oeuvre). The question of when a paraphrase becomes another type of improvisation is difficult to answer; apparently, the dividing line is based purely on intuition. While this answer is somewhat unsatisfactory, it is also inevitable: musicians often reference the melody, to greater or lesser degrees, during improvisation. For example, Hawkins begins his solo on "Epistrophy" by quoting the melody verbatim for eight bars, before developing the rhythmic motives of which the head is composed and producing a varied and fascinating solo. In the present context, melodic renditions that substantially alter the melody will be considered solos or parts of solos, and will

contribute to the evaluation of entropy, while those that do not substantially alter the melody will not be considered part of an excerpt to be evaluated for entropy.

Charlie Parker's rhythmically complex heads were not included in calculations of entropy.

Here I examine several examples of melodic embellishment, namely, embellished versions of "Tea For Two," "All of Me," and" What a Wonderful World," by Lester Young, Doris Day, Ella Fitzgerald, Frank Sinatra, and Louis Armstrong.

"Tea For Two," as written, has very little rhythmic variety. In fact, almost three quarters of the song follow either the template shown in Figure 44 (a), or its close rhythmic relative, Figure 44 (b). The rest of the song consists of whole notes. In all cases, the implied accents are purely metrical, and fall on beats one and three. Not surprisingly, the calculated entropy for this song as written is very low: 0.32, to be exact.

**Figure 44.** Rhythms from which "Tea For Two" is constructed



I examined three renditions of this melody, two by singers – Doris Day and Ella Fitzgerald – and one by Lester Young. Somewhat surprisingly, at least in the case of Doris Day, for whom we might expect lower rhythmic unpredictability than for Fitzgerald or Young, the calculated entropies for these three renditions of the song were quite similar: 2.71 for Fitzgerald, 2.75 for Day, and 2.90 for Young.

"All of Me" has an even lower entropy: every bar begins with an accented note, and accents do not fall on any beats other than downbeats. Thus, the entropy is, in fact, zero! The entropies corresponding to versions by Frank Sinatra, Ella Fitzgerald, and Lester Young, on the other hand, occupy a narrow band of $+/-$ 1.8% centered on $H$ $= 3.54$: Sinatra has $H = 3.477$, Fitzgerald $H = 3.595$, and Young $H = 3.606$. The renditions by Sinatra and Fitzgerald were taken to be the first of two melody choruses, in order to facilitate comparison with the Young melody chorus; had the second melody choruses been used, the entropies most likely would have been greater.

Note, too, that Sinatra employs a subtle behind-the-beat phrasing, not enough to warrant shifting any rhythms by a sixteenth note, as Louis Armstrong is wont to do, e.g. in "What a Wonderful World" m. 23, where beat one is shifted to the second sixteenth note of beat one, but just enough to be perceptible (Figure 45).

Finally, I compare one of Louis Armstrong's 1967 recordings of the aforementioned "What a Wonderful World" (Figure 45) to the sheet music version. I find that, for the sheet music, $H = 1.517$, and for the Armstrong recording, $H = 3.877$.

# Armstrong
## What a Wonderful World



**Figure 45.** Transcription of one of Louis Armstrong's version of "What a Wonderful World"

# CHAPTER 8

## MUTUAL INFORMATION AND SOLOIST-ACCOMPANIST INTERACTION

Here I use the concept of mutual information, as defined in Chapter 2, to analyze the interactions between Charlie Parker and some of the pianists who accompanied him. This is a sensible approach since we expect knowledge of the accented notes in a solo to reflect the placement of chords by the accompanist and vice-versa; mutual information appears to be the perfect tool to represent these interactions.

There are several possible ways to implement the calculation of mutual information for strings of saxophone accents and their corresponding chord onsets. For simplicity, I restrict myself to the case in which IOI's between saxophone accents and between chord onsets are subdivided solely by eighth notes; that is, I omit solos in which sixteenth note, triplet, or straight-eighth-note subdivisions are used in either the soloist or accompanist parts.

Given this restriction, I subdivide the solo accent and chord onset streams into half-measure – that is, four eighth note – groups. Next, I label each group – in both the soloist and accompanist parts – with a number from 1 to 16, according to the 16 possible combinations of four eighth notes or eighth note rests. (There are two possibilities for each of the four positions in each group, thus 2*2*2*2 = 16). I calculate the joint distribution by counting the number of occurrences of each pair of labels (in the soloist and accompanist parts) and dividing by the number of half-measure groups. I calculate the marginal distributions by simply counting the number of occurrences of each label within first the soloist, then the accompaniment parts, and dividing by the number of half-measure groups. Thus I am able to calculate MI scores using equation 2.9.

I begin by considering comping rhythms that result in low MI scores. For an extreme but not far-fetched example, consider comping rhythms with zero entropy. When either distribution in a mutual information calculation has zero entropy, the resulting mutual information is zero as well, since knowledge of one variable has no effect on the knowledge of the other.

Recall that zero entropy rhythms have just one *IOI* value. So, for example, whole notes in the accompaniment part starting on beat one of every measure would yield zero MI, as would quarter notes on every beat. While these rhythms are unlikely to appear in a jazz accompaniment, eighth notes on the "and" of two and the "and" of four can frequently be found in the playing of pianist Red Garland; this, too, would yield zero MI.

Next consider a more complex but still simplistic comping rhythm: the previously discussed "Charleston" rhythm (see Figure 6). Here the MI scores are non-zero, but still low – as compared to the MI scores obtained using the actual comping rhythms (see Table 12), by a factor ranging from 2.63 to 22.65. Thus, as expected, MI reflects to a certain extent the interaction between soloist and accompanist.

Next I calculate MI scores using randomly generated comping rhythms with the same number of chords as in the actual rhythm, and found that, with one exception, they were higher than the scores calculated using the actual comping rhythms. Factors ranged from 0.98 to 2.0. This makes sense since a randomly generated accompaniment would not contain many repeated segments, so knowledge of one's place in the accompaniment would yield a great deal of information about one's place in the solo.

| Title | MI actual | MI Charleston | MI Random | actual/Charleston | random/actual |
|---|---|---|---|---|---|
| KoKo | 0.07 | 0.03 | 0.15 | 2.63 | 2.00 |
| Bloomdido | 0.25 | 0.05 | 0.28 | 5.49 | 1.12 |
| Dewey Square | 0.33 | 0.12 | 0.53 | 2.72 | 1.61 |
| Donna Lee | 0.34 | 0.04 | 0.34 | 9.08 | 0.98 |
| Yardbird Suite | 0.37 | 0.10 | 0.52 | 3.87 | 1.38 |
| Crazeology | 0.40 | 0.09 | 0.43 | 4.46 | 1.08 |
| Cheryl | 0.44 | 0.06 | 0.50 | 7.49 | 1.14 |
| Ornithology | 0.50 | 0.02 | 0.61 | 27.65 | 1.21 |
| Bongo Beep | 0.59 | 0.19 | 0.73 | 3.06 | 1.23 |
| Au Privave | 0.63 | 0.08 | 0.69 | 7.87 | 1.09 |

**Table 12.** Mutual Information for ten Parker excerpts: actual, calculated using a Charleston comping rhythm, and calculated using a random comping rhythm. Ratios of actual MI to MI calculated using Charleston rhythm, MI calculated using random comping rhythm to actual MI.

Finally, I point out a possible reason for the MI score for "KoKo" (recorded 11/26/1945) being lower than the other MI scores. The original pianist booked for this recording session was Bud Powell, but he had to cancel at the last minute; as a replacement, the little known pianist Argonne Thornton (who later changed his name to Sadik Hakim) was brought in. There is disagreement as to whether it was Thornton who played piano on "KoKo," or the trumpeter Dizzy Gillespie filling in on his second instrument. It is unlikely that either of these accompanists would have possessed the pianistic agility displayed by any of the other accompanists represented here (Duke Jordan, Al Haig, Dodo Marmarosa, Bud Powell, Thelonious Monk, or Walter Bishop Jr.). Thus the low MI score; note, too, that the ratio of actual MI to Charleston MI is the lowest of the ten solos evaluated. So this is an important result.

# CHAPTER 9

## CONCLUSION

Within roughly ten years of its founding as a discipline, information theory had been applied to music by several authors. One reason for this is surely that the formula for Shannon entropy – the central concept of information theory – is simple, and can be used with any random variable described by a known probability distribution. Another, more enlightening, reason is provided by Meyer (1957), who posits that, in music, the phenomenon of expectation, specifically *thwarted* expectation, correlates with both information and musical meaning.

The present work explored the relationship of entropy, and of the related concept of mutual information, to jazz rhythm in several ways. It made a brief foray into the study of rhythmic periodicity and syncopation as they pertain to the perception of rhythmic complexity. And it used stress and contour accent density to provide a possible explanation for counterintuitive results.

First of all, it demonstrated that entropy derived from IOIs between accented notes in jazz solos is a significant factor in the perception of rhythmic complexity for jazz rhythms constructed solely of multiples of eighth-notes and eighth-note rests. This conclusion was reached by means of an experiment that asked fifteen music majors to rank eighteen short rhythmic excerpts for complexity on a scale from one to seven. Results were analyzed using a mixed-effects multiple regression model with the Satterthwaite approximation. Multiple predictor variables were used, in addition to entropy: number of notes (which was roughly the same as note density because excerpts were approximately the same length), several variables quantifying the periodicity of test

rhythms (or the lack thereof), syncopation, order effects (in other words, the effects of listening to the experimental excerpts in different orders), and jazz experience level. In the course of identifying models that include entropy, number of notes, and excerpt order as significant factors, I also found models that include lack of periodicity, syncopation, number of notes, and order effects as significant factors in the perception of rhythmic complexity. It is likely that entropy was mediated by periodicity in its effect on rhythmic complexity ratings.

Periodicity reflects meter, since rhythms that are tethered to the underlying meter will have high periodicity; syncopation, too, reflects the interaction of rhythm and meter. Entropy, however, has no dependence on meter, dependent as it is on inter-onset intervals regardless of the placement of notes in relation to meter.

Next, using a corpus of 88 transcribed solos by Louis Armstrong, Coleman Hawkins, Lester Young, Charlie Christian, and Charlie Parker, I demonstrated that solos by Young are more entropic than those by Armstrong, Christian, and Parker, but not Hawkins. I arrived at this conclusion by isolating dynamical accented notes and calculating probability distributions based on the inter-onset intervals between accented notes, calculating the entropy of each excerpt, and using the estimated marginal means technique with the Bonferroni correction to make pairwise comparisons between musicians. Two covariates were used in this calculation: number of distinct IOIs and number of accents. The utility of these choices was confirmed using the added variable plot technique; including these covariates obviated the need for normalizing the computed entropies.

The lack of difference in entropy between Hawkins and Parker was unexpected, since the received wisdom regarding these two musicians is that Hawkins is more predictable, rhythmically speaking, than Parker. This anomaly is probably explained by examining dynamical accent density and contour accent density. Hawkins is found to have significantly greater dynamical accent density, while Parker is found to have significantly greater contour accent density. These results are in keeping with the received wisdom.

The phenomenon of melodic embellishment, central to jazz, was explored using entropy. Embellished versions of standard songs were uniformly found to have higher entropies than their sheet-music counterpoints.

Mutual information was used to study the interaction between Parker and his piano accompanists in a limited corpus of ten transcribed solos. Mutual information between Parker and his accompanists was found to be greater than that between Parker and a repeating Charleston rhythm accompaniment, and less than that between Parker and a random accompaniment. The latter result makes sense because knowledge of one's place in a random accompaniment yields a high amount of information about one's place in the solo. The lowest mutual information value and lowest ratio of mutual information between Parker and his accompanist to the mutual information between Parker and a Charleston accompaniment occurs on "KoKo"; this may be due to the fact that the accompanist on the date, either Argonne Thornton (Sadik Hakim) or Dizzy Gillespie, was probably less pianistically agile than Parker's usual accompanists.

Thus, this research has indicated that entropy is probably a significant factor in the perception of rhythmic complexity, that entropy depends on musician in a corpus of

128

88 solos by five great jazz musicians, that entropy is a useful tool for understanding melodic embellishment, and that mutual information is a useful tool for studying the interaction between soloist and accompanist.

It points the way toward at least four avenues for future exploration.

First and foremost, this research calls for experimental studies of entropy as a significant factor in the perception of rhythmic complexity for a more general class of rhythms, those including, triplets, sixteenth notes, and ornamental straight eighths. It is possible that the conclusions reached in the present study might be changed in a more general experimental context. For example, the presence or absence of triplets, sixteenth notes and ornamental straight eighths might overwhelm the effects of entropy on perceived rhythmic complexity. Or the effects of number or density of notes on complexity might interact with the types of rhythm included: perhaps the number of quarter-note triplets might have a stronger effect on perceived rhythmic complexity than the number of eighth-notes alone.

This research calls for studying carryover effects in such a way as to isolate the effects of predictor variables on subsequent complexity ratings. It seems intuitive that an excerpt with high entropy followed by an excerpt with low entropy might artificially deflate the subject's reaction to the second excerpt. It is unlikely, however, that this would significantly change the experimental results obtained here, if the negligible magnitude of the carryover effects isolated here is any indication.

It calls for computational research with more excerpts and more musicians. An equal number of excerpts should be transcribed for all musicians, and this should be as

high as possible. And adding other musicians to the solos might help identify what exactly the factors are that lead to low or high entropy as it was here defined. Would Lester Young's solos be more or less entropic than John Coltrane's? Or Benny Goodman's? An alternative to transcribing more solos would be to explore extant transcriptions of jazz solos available online. This might be an efficient way to expand the corpus.

And finally, it calls for corpora that are larger and include musicians other than only Charlie Parker in the study of mutual information. The evidence is, however, that mutual information is a useful tool in studying soloist/accompanist interaction. Perhaps there are specific issues that can be explored using mutual information, similar to the anecdote regarding Sadik Hakim/Dizzy Gillespie.

Thus, building on previous work in the field of information theory as applied to classical and jazz music, and using both experimental and computational techniques, this work adds a valuable new perspective on the study of rhythmic complexity in jazz.

# APPENDIX A

# EXPERIMENTAL EXCERPTS

EXCERPT 1.1



EXCERPT 1.2

EXCERPT 2.1



EXCERPT 2.2

EXCERPT 3.1



EXCERPT 3.2

EXCERPT 4.1



EXCERPT 4.2

EXCERPT 5.1



EXCERPT 5.2

## EXCERPT 6.1



## EXCERPT 6.2

EXCERPT 7.1



EXCERPT 7.2

EXCERPT 8.1



EXCERPT 8.2

EXCERPT 9.1



EXCERPT 9.2

## APPENDIX B

## CATALOG OF EXCERPTS

| Name | Date | Title | TS | IOIs | acc's | Entropy |
|---|---|---|---|---|---|---|
| Armstrong | 26 | Cornet Chop Suey pt. 2 | | 14 | 84 | 3.138 |
| Armstrong | 27 | Potato Head Blues pt. 1 | | 12 | 31 | 3.24 |
| Armstrong | 27 | Potato Head Blues pt. 2 | | 12 | 100 | 3.03 |
| Armstrong | 29 | Mahogany Stomp pt. 1 | | 8 | 19 | 2.795 |
| Armstrong | 29 | Mahogany Stomp pt. 2 | | 8 | 14 | 2.815 |
| Armstrong | 29 | Mahogany Stomp pt. 3 | | 10 | 54 | 2.58 |
| Armstrong | 31 | Stardust pt. 1 | ✓ | 15 | 39 | 3.491 |
| Armstrong | 31 | Stardust pt. 2 | ✓ | 18 | 45 | 3.327 |
| Armstrong | 31 | Stardust pt. 3 | ✓ | 23 | 55 | 4.174 |
| Armstrong | 38 | I Double Dare You | | 15 | 65 | 3.11 |
| Armstrong | 43 | I Never Knew | ✓ | 12-13 | 33-35 | 3.201 |
| Armstrong | 44 | I'm Confessin' (That I Love You) | | 14 | 43 | 3.208 |
| Armstrong | 47 | It Takes Time | ✓ | 9 | 36 | 2.785 |
| Armstrong | 53 | The Gypsy | ✓ | 19 | 71 | 3.305 |
| Armstrong | 56 | A Foggy Day | ✓ | 11 | 32 | 3.02 |
| Armstrong | 63 | Hello Dolly | | 13 | 39 | 2.803 |
| Armstrong | 67 | What a Wonderful World | ✓ | 23 | 79 | 3.877 |
| Hawkins | 26 | The Stampede | | 11 | 72 | 2.9 |
| Hawkins | 33 | Talk of the Town | | 16 | 76 | 3.393 |
| Hawkins | 39 | Body and Soul | ✓ | 20 | 236 | 2.773 |
| Hawkins | 40 | Bouncin' With Bean | | 12 | 86 | 3.027 |
| Hawkins | 40 | My Blue Heaven | | 9 | 75 | 2.046 |
| Hawkins | 41 | Disorder at the Border | | 12 | 108 | 2.724 |
| Hawkins | 44 | Flyin' Hawk | ✓ | 15 | 100 | 3.336 |
| Hawkins | 50 | Ballade pt. 1 | ✓ | 15 | 57 | 3.238 |
| Hawkins | 50 | Ballade pt. 2 | ✓ | 15 | 55 | 3.435 |
| Hawkins | 54 | Lullaby of Birdland | ✓ | 16 | 55 | 3.341 |
| Hawkins | 57 | Blues for Tomorrow | ✓ | 17 | 215 | 2.002 |
| Hawkins | 57 | Epistrophy | | 18 | 82 | 3.665 |
| Hawkins | 62 | Satin Doll | ✓ | 24 | 215 | 2.862 |
| Hawkins | 62 | Wanderlust | ✓ | 21 | 62 | 3.794 |
| Hawkins | 63 | Just Friends | | 15 | 79 | 3.122 |
| Young | 36 | Lady Be Good | | 16 | 99 | 3.465 |
| Young | 36 | Shoe Shine Boy | | 20 | 86 | 3.656 |
| Young | 38 | Back in Your Own Backyard | | 13 | 51 | 3.271 |
| Young | 38 | When You're Smiling | | 12 | 44 | 3.071 |
| Young | 42 | Body and Soul pt. 1 | ✓ | 19 | 73 | 3.616 |
| Young | 42 | Body and Soul pt. 2 | ✓ | 14 | 52 | 3.392 |
| Young | 42 | Indiana pt. 1 | ✓ | 18 | 92 | 3.798 |
| Young | 42 | Indiana pt. 2 | | 21 | 92 | 3.557 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Young | 42 | Tea For Two 1942 | | 17 | 113 | 3.189 |
| Young | 45 | D.B. Blues | ✓ | 17 | 87 | 3.64 |
| Young | 46 | It's Only a Paper Moon | | 21 | 163 | 3.328 |
| Young | 47 | Sheik of Araby pt. 1 | | 17 | 77 | 3.624 |
| Young | 47 | Sheik of Araby pt. 2 | | 14 | 46 | 3.463 |
| Young | 47 | Tea For Two 1947 | | 17 | 82 | 3.435 |
| Young | 50 | Blues For Greasy JATP | | 21 | 38 | 4.186 |
| Young | 52 | Ad Lib Blues pt. 1 | | 22 | 132 | 3.441 |
| Young | 52 | Ad Lib Blues pt. 2 | | 19 | 194 | 3.158 |
| Young | 56 | All of Me pt. 2 | | 21 | 117 | 3.692 |
| Young | 56 | Taking a Chance on Love | ✓ | 22 | 134 | 3.822 |
| Young | 56 | You Can Depend on Me | ✓ | 22 | 158 | 3.658 |
| Christian | 39 | Christian Honeysuckle Rose | | 13 | 64 | 2.83 |
| Christian | 39 | Flying Home | | 10 | 61 | 2.789 |
| Christian | 39 | Good Morning Blues | | 17 | 46 | 3.497 |
| Christian | 39 | Rose Room | ✓ | 17 | 60 | 3.043 |
| Christian | 39 | Seven Come Eleven | ✓ | 17 | 43 | 3.638 |
| Christian | 40 | Benny's Bugle | | 13 | 35 | 3.241 |
| Christian | 40 | Gone With "What" Wind | | 13 | 39 | 3.22 |
| Christian | 40 | Grand Slam | | 16 | 36 | 3.557 |
| Christian | 40 | I Can't Give You Anything But… | ✓ | 13 | 34 | 3.002 |
| Christian | 40 | Six Appeal | ✓ | 9 | 36 | 2.695 |
| Christian | 40 | Till Tom Special | | 11 | 20 | 3.003 |
| Christian | 40 | Wholly Cats | | 11 | 33 | 3.097 |
| Christian | 41 | Breakfast Feud | | 13 | 38 | 3.199 |
| Christian | 41 | I've Found a New Baby | | 10 | 46 | 2.828 |
| Parker | 40 | Moten Swing | | 13 | 58 | 2.522 |
| Parker | 40 | Parker Honeysuckle Rose | | 10 | 62 | 2.199 |
| Parker | 40 | Parker Lady Be Good | | 12 | 50 | 2.609 |
| Parker | 41 | Swingmatism | | 11 | 18 | 3.264 |
| Parker | 42 | Hootie's Blues | ✓ | 12 | 32 | 3.011 |
| Parker | 45 | Billie's Bounce | | 17 | 66 | 3.351 |
| Parker | 46 | Yardbird Suite | | 14 | 48 | 3.556 |
| Parker | 47 | Bongo Beep | ✓ | 13 | 40 | 3.216 |
| Parker | 47 | Cheryl | | 18 | 46 | 3.538 |
| Parker | 47 | Crazeology | | 12 | 53 | 2.699 |
| Parker | 47 | Ornithology | | 17 | 43 | 3.461 |
| Parker | 48 | Segment | | 17 | 93 | 3.414 |
| Parker | 49 | Scrapple From the Apple | | 17 | 42 | 3.747 |
| Parker | 50 | Bloomdido | | 17 | 73 | 2.972 |
| Parker | 50 | Mohawk | | 20 | 78 | 3.656 |
| Parker | 51 | Au Privave | | 19 | 44 | 3.952 |
| Parker | 51 | Blues For Alice | | 21 | 54 | 4.033 |
| Parker | 51 | She Rote | ✓ | 22 | 93 | 3.554 |
| Parker | 53 | Chi Chi | ✓ | 21 | 102 | 3.794 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Parker | | 53 | Now's The Time | ✓ | 25 | 78 | 4.157 |
| Parker | ?? | | Dewey Square | | 16 | 46 | 3.297 |
| Parker | ?? | | Donna Lee | | 21 | 86 | 3.756 |

*TS checked indicates a 2:1 swing ratio and dependence of entropy on swing ratio*

*For Armstrong/I Never Knew, ranges of num_IOI and num_accent, and average entropy, are given for different dispositions of Armstrong's "clams"*

*For a detailed list of recording data, please contact the author at dougabrams.jazz@gmail.com*

# APPENDIX C

## TREATMENT OF ESTIMATED ERRORS

There were 6,299 accented notes included in the corpus, and thus easily 20,000–30,000 notes altogether. In a manually transcribed corpus of this size, there are bound to be errors. Overall, the error rate – estimated using a random sample of 37 excerpts – was fairly low: insertion, deletion, or translation of accented notes as a percentage of number of accented notes was about 0.5%, and as a percentage of all notes, was lower. There is a potential problem, however, in that the errors are distributed evenly across excerpts; thus, the number of excerpts with minor errors as a percentage of the total number of excerpts – estimated using the same random sample – was much higher: about 40%. Given that files discovered with errors were corrected, this means that in the corpus as a whole, there were probably about 20 files with minor errors after review and correction.

Once errors were discovered, I used four methods of quantifying them. Percentage errors in entropy, number of IOIs, and number of accents were identified, and pairwise *p* values were calculated before and after correction. For  a limited corpus of twelve error–containing excerpts, Table 13 gives the mean, median, and standard deviation of percentage error in entropy, percentage error in number of IOIs, and percentage error in number of accents. Note that the number of IOIs exhibits the largest average percent error, since making a small change to the data can result in a change in number of IOIs of one or even two, both of which are relatively large in terms of number of IOIs. Finally, including the twelve error-containing excerpts mentioned above and correcting them one by one did not qualitatively change the pairwise *p* values.

While *p* values are relatively insensitive to simple errors, a review of the data indicates that adding or deleting excerpts for various reasons has a more pronounced effect. In the course of analyzing the data, I realized that some excerpts had to be deleted for reasons having to do with the properties of the swing eighth-notes used, either because they were too close to straight-eighths, or because it was too difficult to ascertain whether or not the eighth-note pairs had a 2:1 ratio. For example, for a data set with five error-containing excerpts and with just one more excerpt than the final data set, the Christian/Young *p* value was 0.0406, while for the same data set with *four* more excerpts than the final data set, the *p* value was 0.0603. Furthermore, a data set from early in the experiment with many more error-containing excerpts (27) *and* with four more excerpts than the final data set yielded a *p* value for the Christian/Young pairwise comparison of 0.0541. For all data sets examined, the only *p* values that differ qualitatively from those obtained from the final set are those corresponding to the Christian/Young comparison, and these were at least "marginally significant" (0.0518–0.0627). In any case, it appears as though adding or deleting excerpts has a greater effect on *p* values than the presence or absence of simple errors.

Having settled on a final corpus, and estimating that no more than twenty excerpts contain errors, I argue that the *p* values for Christian/Young and for all other pairwise comparisons are valid. Even if there are twenty error-containing excerpts, it appears that correcting errors drives the Christian/Young *p* value down, therefore not affecting significance. There is, however, a chance that the presence of unknown errors may have an unpredictable effect on the Christian/Young *p* value, perhaps resulting in marginal

144

significance or even insignificance. In this case, the Armstrong/Young and Parker/Young

*p* values still indicate significant pairwise comparisons.

| Quantity/Statistic | Mean | Median | Standard Deviation |
|---|---|---|---|
| Entropy | 1.603 | 1.1 | 1.213 |
| Number of IOIs | 3.683 | 4.7 | 3.510 |
| Number of accents | 0.75 | 0.95 | 0.722 |

**Table 13.** Statistical descriptions of three measures of percentage error: Entropy, Number of IOIs, and Number of accents for a corpus of 12 error-containing excerpts

# REFERENCES

Belgrad, Daniel. *The culture of spontaneity: Improvisation and the arts in postwar America*. University of Chicago Press, 1998.

Bell, Chip. "Algorithmic music composition using dynamic Markov chains and genetic algorithms." *Journal of Computing Sciences in Colleges* 27, no. 2 (2011): 99–107.

Blesh, Rudi, and Harriet Janis. *They All Played Ragtime–The True Story of an American Music*. Read Books Ltd, 2011.

Carrabré, Ariel. "Understanding Schenkerian Analysis from the Perspective of Music Perception and Cognition." PhD diss., Université d'Ottawa/University of Ottawa, 2015.

Chilton, John. *The song of the Hawk: The life and recordings of Coleman Hawkins*. University of Michigan Press, 1990.

Christensen, Thomas, ed. *The Cambridge history of Western music theory*. Cambridge University Press, 2010

Cohen, Joel E. "Information theory and music." *Behavioral Science* 7, no. 2 (1962): 137–163.

Coons, Edgar, and David Kraehenbuehl. "Information as a measure of structure in music." *Journal of Music Theory* 2, no. 2 (1958): 127–161.

Cover, Thomas M., and Joy A. Thomas. "Elements of information theory 2nd edition (Wiley series in telecommunications and signal processing)." (2006).

Daikoku, Tatsuya. "Statistical Properties in Jazz Improvisation Underline Individuality of Musical Representation." *NeuroSci* 1, no. 1 (2020): 4.

De Fleurian, Remi, Oded Ben-Tal, Daniel Müllensiefen, and Tim Blackwell. "Comparing perceptual and computational complexity for short rhythmic patterns." *Procedia-Social and Behavioral Sciences* 126 (2014): 111-112.

DeFleurian, Remi, Tim Blackwell, Oded Ben-Tal, and Daniel Müllensiefen. "Information-theoretic measures predict the human judgment of rhythm complexity." *Cognitive Science* 41, no. 3 (2017): 800-813.

DeSouza, Jonathan. "Melodic Transformation in George Garzone's Triadic Chromatic Approach; or, Jazz, Math, and Basket Weaving." *Music Theory Spectrum* (2022).

Deutsch, Diana. "Octave generalization and tune recognition." *Perception & Psychophysics* 11, no. 6 (1972): 411-412.

DeVeaux, Scott. *The birth of bebop*. University of California Press, 1997.

Devore, Jay L., and Kenneth N. Berk. *Modern mathematical statistics with applications*. New York: Springer, 2012.

Dibben, Nicola. "The cognitive reality of hierarchic structure in tonal and atonal music." *Music Perception* 12, no. 1 (1994): 1-25.

Dowling, W. Jay. "Scale and contour: Two components of a theory of memory for melodies." *Psychological review* 85, no. 4 (1978): 341.

Duane, Ben. "Agency and Information Content in Eighteenth and Early Nineteenth–Century String–Quartet Expositions." *Journal of Music Theory* 56, no. 1 (2012): 87–120.

Edworthy, Judy. "Interval and contour in melody processing." *Music Perception* 2, no. 3 (1985): 375-388.

Ellis, Robert J., and Mari R. Jones. "The role of accent salience and joint accent structure in meter perception." *Journal of Experimental Psychology: Human Perception and Performance* 35, no. 1 (2009): 264.

Fitch, W. Tecumseh, and Andrew J. Rosenfeld. "Perception and production of syncopated rhythms." *Music Perception* 25, no. 1 (2007): 43–58.

Franz, David Matthew. "Markov chains as tools for jazz improvisation analysis." PhD diss., Virginia Tech, 1998.

Friedmann, Michael L. "A methodology for the discussion of contour: Its application to Schoenberg's music." *Journal of Music Theory* 29, no. 2 (1985): 223-248.

Frieler, Klaus. "Constructing Jazz Lines Taxonomy, Vocabulary, Grammar." *Jazzforschung heute: Themen, Methoden, Perspektiven* (2019): 103–132.

Gillick, Jonathan Reuven. "A clustering algorithm for recombinant jazz improvisations." *Honors Theses* (2009): 345.

Gillick, Jon, Kevin Tang, and Robert M. Keller. "Learning jazz grammars." In *Proceedings of the sound and music computing conference*, pp. 125–130. 2009.

Gillick, Jon, Kevin Tang, and Robert M. Keller. "Machine learning of jazz grammars." *Computer Music Journal* 34, no. 3 (2010): 56–66.

Hiller, Lejaren, and Ramon Fuller. "Structure and information in Webern's Symphonie, Op. 21." *Journal of Music Theory* 11, no. 1 (1967): 60–115.

Huron, David. *Sweet anticipation: Music and the psychology of expectation*. MIT press, 2008.

Iyer, Vijay. "Embodied mind, situated cognition, and expressive microtiming in African–American music." *Music perception* 19, no. 3 (2002): 387–414.

Jacoby, Nori, Naftali Tishby, and Dmitri Tymoczko. "An information theoretic approach to chord categorization and functional harmony." *Journal of New Music Research* 44, no. 3 (2015): 219–244.

Joel, Lester. *The Rhythms of Tonal Music*. Carbondale: University of Illinois Press. 1986

Jones, LeRoi. *Blues people: Negro music in white America*. Perennial, 2002.

Keil, Charles, and Steven Feld. "Music Grooves Essays and Dialogues." (1994).

Kernfeld, Barry. ed. 1996. *New Grove Dictionary of Jazz*. New York: St. Martin's Press

Knopoff, Leon, and William Hutchinson. "Entropy as a measure of style: The influence of sample length." *Journal of Music Theory* 27, no. 1 (1983): 75–97.

Knopoff, Leon, and William Hutchinson. "Information theory for musical continua." *Journal of Music Theory* 25, no. 1 (1981): 17–44.

Krämer, Ulrich. "SCHOENBERG'S CONCEPTS OF" KOMPOSITIONSLEHRE"(1904-1911) AND THE NINETEENTH-CENTURY GERMAN TRADITION." *Revista de Musicología* 16, no. 6 (1993): 3735-3753.

Krumhansl, Carol L., and Edward J. Kessler. "Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys." *Psychological review* 89, no. 4 (1982): 334.

Leong, Daphne. "Generalizing syncopation: Contour, duration, and weight." *Theory and Practice* (2011): 111–150.

Linskens, Erlijn J., and Gijs Schoenmakers. "Music Improvisation Using Markov Chains." *Maastricht University* (2014).

Longuet–Higgins, H. Christopher, and Christopher S. Lee. "The rhythmic interpretation of monophonic music." *Music Perception* 1, no. 4 (1984): 424–441.

Margulis, Elizabeth Hellmuth, and Andrew P. Beatty. "Musical style, psychoaesthetics, and prospects for entropy as an analytic tool." *Computer Music Journal* 32, no. 4 (2008): 64–78.

Mayhew, Jonathan. 2022. "Improvisation." *Arcade* (blog), October 28. https://shc.stanford.edu/arcade/interventions/improvisation

Meyer, Leonard B. "Meaning in music and information theory." *The Journal of Aesthetics and Art Criticism* 15, no. 4 (1957): 412–424.

Morris, Robert D. "New directions in the theory and analysis of musical contour." *Music Theory Spectrum* 15, no. 2 (1993): 205-228.

Norgaard, Martin, Jonathan Spencer, and Mariana Montiel. "Testing cognitive theories by creating a pattern–based probabilistic algorithm for melody and rhythm in jazz improvisation." *Psychomusicology: Music, Mind, and Brain* 23, no. 4 (2013): 243.

Parncutt, Richard. "A perceptual model of pulse salience and metrical accent in musical rhythms." *Music perception* 11, no. 4 (1994): 409–464.

Patel, Aniruddh D., and Joseph R. Daniele. "An empirical comparison of rhythm in language and music." *Cognition* 87, no. 1 (2003): B35–B45.

Pearce, Marcus, and Geraint Wiggins. "Improved methods for statistical modelling of monophonic music." *Journal of New Music Research* 33, no. 4 (2004): 367–385.

Pfleiderer, Martin, and Klaus Frieler. "The Jazzomat project. Issues and methods for the automatic analysis of jazz improvisations." *Concepts, experiments, and fieldwork: Studies in systematic musicology and ethnomusicology* (2010): 279–295.

Pfordresher, Peter Q. "The role of melodic and rhythmic accents in musical structure." *Music Perception* 20, no. 4 (2003): 431-464.

Jones, Mari Riess, and Peter Q. Pfordresher. "Tracking musical patterns using joint accent structure." *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale* 51, no. 4 (1997): 271.

Parker, Charlie, and Jamey Aebersold. *Charlie Parker Omnibook: for C instruments (treble clef)*. Alfred Music Publishing, 1978.

Pinkerton, Richard C. "Information theory and melody." *Scientific American* 194, no. 2 (1956): 77–87.

Polansky, Larry. "Morphological metrics." *Journal of New Music Research* 25, no. 4 (1996): 289-368.

Raphael, Christopher, and Joshua Stoddard. "Functional harmonic analysis using probabilistic models." *Computer Music Journal* 28, no. 3 (2004): 45–52.

Quick, Donya, and Kelland Thomas. "A functional model of jazz improvisation." In *Proceedings of the 7th ACM SIGPLAN International Workshop on Functional Art, Music, Modeling, and Design*, pp. 11–21. 2019.

Quinn, Ian. "The combinatorial model of pitch contour." *Music Perception* 16, no. 4 (1999): 439-456.

Roeder, John. "A calculus of accent." *Journal of Music Theory* (1995): 1-46.
Rouse, Tevin. "Stochastic Improvisation of Jazz Solos." (2017).

Sadakata, Makiko, Peter Desain, and Henkjan Honing. "The Bayesian way to relate rhythm perception and production." *Music Perception* 23, no. 3 (2006): 269–288.

Shannon, Claude E. "A mathematical theory of communication." *The Bell system technical journal* 27, no. 3 (1948): 379–423.

Schuller, Gunther. *Early jazz: Its roots and musical development*. Vol. 1. History of Jazz. New York and Oxford: Oxford University Press, 1968.

Schuller, Gunther, and Martin Williams. "Liner notes to Big Band Jazz: From the Beginnings to the Fifties." *Smithsonian RD* 30 (1983).

Schmuckler, Mark A. "Testing models of melodic contour similarity." *Music Perception* 16, no. 3 (1999): 295-326.

Serafine, Mary Louise, Noah Glassman, and Cornell Overbeeke. "The cognitive reality of hierarchic structure in music." *Music Perception* 6, no. 4 (1989): 397-430.

Shmulevich, Ilya, and Dirk–Jan Povel. "Complexity measures of musical rhythms." *Rhythm perception and production* (2000): 239–244.

Smith, Leigh M., and Henkjan Honing. "Evaluating and extending computational models of rhythmic syncopation in music." In *ICMC*. 2006.

Snyder, John L. "Entropy as a measure of musical: The influence of a priori assumptions." *Music Theory Spectrum* 12, no. 1 (1990): 121–160.

Tan, Ivan, Ethan Lustig, and David Temperley. "Anticipatory syncopation in rock: A corpus study." *Music Perception: An Interdisciplinary Journal* 36, no. 4 (2019): 353–370.

Temperley, David. "Syncopation in rock: a perceptual perspective." *Popular music* 18, no. 1 (1999): 19–40. style

Temperley, David. "An evaluation system for metrical models." *Computer Music Journal* 28, no. 3 (2004): 28–44.

Temperley, David. *Music and probability*. MIT Press, 2007.

Temperley, David. "Modeling common–practice rhythm." *Music Perception* 27, no. 5 (2010): 355–376.

Temperley, David. "Rhythmic variability in European vocal music." *Music Perception: An Interdisciplinary Journal* 35, no. 2 (2017): 193–199.

Temperley, David. "Second–position syncopation in European and American vocal music." *Empirical Musicology Review* 14, no. 1–2 (2019): 66–80.

Temperley, David. "Uniform information density in music." *Music Theory Online* 25, no. 2 (2019).

Thul, Eric. "Measuring the complexity of musical rhythm." (2008)

Thul, Eric, and Godfried T. Toussaint. "Rhythm Complexity Measures: A Comparison of Mathematical Models of Human Perception and Performance." In *ISMIR*, pp. 663-668. 2008.

Toussaint, Godfried T. "The pairwise variability index as a tool in musical rhythm analysis." In *Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Science of Music*, pp. 23-07. 2012.

Toussaint, Godfried T. "The pairwise variability index as a measure of rhythm complexity." *Analytical Approaches to World Music* 2, no. 2 (2013): 1-42.

Victor, Joseph. "Jazz Melody Generation and Recognition." (2012).

Vitz, Paul C. "Information, run structure and binary pattern complexity." *Perception & Psychophysics* 3, no. 4 (1968): 275-280.

Vitz, Paul C., and Thomas C. Todd. "A coded element model of the perceptual processing of sequential stimuli." *Psychological Review* 76, no. 5 (1969): 433.

Wang, Cheng–I., and Shlomo Dubnov. "Context–aware hidden Markov models of jazz music with variable Markov oracle." In *Proceedings of 5th international workshop on musical Metacreation (MUME2017), Atlanta, USA*, pp. 19–23. 2017.

Weisberg, Sanford. *Applied linear regression*. Vol. 528. John Wiley & Sons, 2005.

Williams, Martin "Williams, Martin. *The Smithsonian collection of classic jazz*. Smithsonian Collection of Recordings, 1997.

Youngblood, Joseph E. "Style as information." *Journal of Music Theory* 2, no. 1 (1958): 24–35.

Yun, Kristy. "Constraints and Markov chains for a Cognitive Model of Jazz Improvisation: Adding Chords." (2016).

Yust, Jason. "Testing Schenkerian theory: An experiment on the perception of key distances." In *Proceedings of the 2012 International Conference for Music Perception and Cognition*. 2012.