nothing to do with nonlinear coupling of modes. Since the FT cannot discriminate between these effects, we need better methods.

ACKNOWLEDGMENTS

I would like to thank the Santa Fe Institute and the organizers of the time series workshop for encouraging interaction between the experts who develop new time series analysis techniques and those who can benefit by applying them in their various disciplines.

†Department of Music, Stanford University, Stanford, CA 94305;
E-mail: mdirst@leland.stanford.edu
‡Xerox PARC, 3333 Coyote Hill Road, Palo Alto, CA 94304;
Address for correspondence: Andreas Weigend, Department of Computer Science and Institute of Cognitive Science, University of Colorado, Boulder, CO 80309-0430;
E-mail: weigend@cs.colorado.edu

Baroque Forecasting: On Completing J. S. Bach's Last Fugue

An unfinished work of art is eternally provocative. Unfinished musical works, particularly those by important composers, are rarely left incomplete: midwife musicians learn to recreate the style of a composer in order to finish them. The recent availability of musical texts in machine-readable form allows us to apply methods of statistics, machine learning, and artificial intelligence to the formerly exclusive domain of historically minded composers and musicologists. The scientific approach, while it may begin with the same goal as traditional inquiry—i.e., How do we finish the piece?—leads to new questions and points in new directions.

To provide a testing ground for these new questions and methods, the organizers of the Santa Fe Time Series Analysis and Prediction Competition selected one of the most enigmatic unfinished works in music history: J. S. Bach's last fugue, Contrapunctus XIV from *Die Kunst der Fuge*.

We address three different tasks: analysis, continuation, and completion. While we make no attempt to actually complete the fugue, we apply statistical methods in order to characterize the data set, and we relate the

Times Series Prediction: Forecasting the Future and Understanding the Past, Eds. A. S. Weigend and N. A. Gershenfeld, SFI Studies in the Sciences of Complexity, Proc. Vol. XV, Addison-Wesley, 1993

musical text to perception and cognition. We emphasize the importance of hierarchical structure and discuss the effects of different representations.

We contrast this "data-driven" approach (in which features are learned from the data) to rule-based expert systems, and to various completions by composers and musicologists. Finally, by reexamining Bach's manuscript, we add a new twist to the detective story of Contrapunctus XIV.

1. INTRODUCTION

In modern time series analysis and prediction, one of the recurring themes is the tension between randomness and order, between stochastic and deterministic models. The realization that simple functions can produce sequences that look complex and even pass standard tests for randomness has had an impact on time series prediction. Unfortunately, it is sometimes difficult to separate the hope from the hype. Our hope in this paper is to see whether this fact—that simple nonlinear models can generate apparently complex behavior—has any relevance for music. To clarify our scope, we are not addressing issues of sound (such as timbre), performance (such as articulation), or instrumentation: we focus on the notes.

Musical examples of this tension between randomness and order range from stochastic instruments—e.g., wind-driven chimes and Aeolian harps (the use of which nearly cost Saint Dunstan (d. 988) his life for suspected sorcery)—to deterministic compositional techniques. Some examples of the latter include: Johannes Kepler's (1619) calculation of melodies based on the orbits of the planets, the use of Bach's name as a melodic figure (the notes B.A.C.H. in German notation correspond to B♭.A.C.B♮. in English notation), or the composition of a melody inspired by images, such as the skyline of San Francisco or Hong Kong, or a bunch of bent nails strewn on the ground.[1] Mixing randomness and order, Samuel Pepys (1639–1703) used decks of cards to "draw" melodic tunes, and W. A. Mozart (1787) constructed algorithms for the random combination of subsequences and called them *Musikalisches Würfelspiel*, a musical game of dice.[2]

The tension between randomness and order is also important in music perception. Our perception of music is controlled by expectations, which are generated by music's deterministic structure. In principle, deterministic structure—the regularities—can be extracted by using artificial intelligence techniques. The more

[1]Loy (1991) mentions that this technique of generating melodies by casting bent nails on the ground (suggested by Vogt in Prague around 1719) served mainly "to prime the pump, so to speak, of a composer's imagination," and was not intended to be completely deterministic.

[2]Two examples are KV Anh. 294d and KV 516f, reprinted in Cope (1991) and in Schwanauer and Levitt (1993), respectively.

TABLE 1 Transformations of the fugue subject.

musical term	operation		
transposition[1]	$x \leftarrow x + c$	(translation)	move to a different pitch level
retrograde	$t \leftarrow -t$	(time reversal)	play backward
inversion	$d \leftarrow -d$	(pitch reflection)	play mirror image
diminution	$t \leftarrow 2t$		play twice as fast
augmentation	$t \leftarrow 0.5t$		play twice as slow

[1] The operation of *transposition* sometimes requires slight changes in the intervallic structure of the theme. For example, in the opening measures of a fugue, where each voice enters in turn, the theme is sometimes adjusted in order to remain within the key (the tonality) of the fugue. This type of alteration produces a *tonal* answer; a *real* answer replicates exactly the intervallic structure of the subject.

structure in a piece of music, the higher the chance that a machine learning approach will succeed. The organizers of the Competition selected a fugue, because it has a high amount of structure, certainly more than the foreign exchange rate data set of the Competition.

A **fugue** typically has one primary *theme* (a fugue *subject*) and may or may not have secondary themes (*countersubjects*). These themes are processed by a number of symmetry transformations, shown in Table 1.

A good fugue uses the theme(s) in all the voices, combining the theme(s) with transformations as often, and in as many artful ways, as possible. But a random combination of the theme and its modified versions in different keys is hardly the essence of a fugue. There are many constraints: thematic and nonthematic material must fit together for a musical work to make harmonic and rhythmic sense. Composing a fugue can be viewed as an optimization problem.

The music of Johann Sebastian Bach (1685–1750) contains an embarrassment of riches. His *Kunst der Fuge*, BWV 1080 (hereafter KdF), a multimovement summation of the fugal art, includes 14 *Contrapuncti* (fugues), 2 inversions of these fugues, and 4 canons—all of which are related by the use of a single theme, the *KdF theme*, which during the course of the work undergoes subtle variations in rhythm and melody that serve to distinguish the individual pieces.

Contrapunctus XIV from Bach's *Art of Fugue* (hereafter abbreviated Cp. XIV)[3] is incomplete, making it a prime candidate for analysis and attempted continuation and completion.[4] Although Cp. XIV, as it stands in Bach's manuscript, does not contain the KdF theme, it has long been acknowledged as part of KdF.[5] The issue of *why* Cp. XIV is incomplete will be addressed from a musicological perspective in Section 4. (The first and last pages of Bach's manuscript of Cp. XIV are reproduced in Figure 1.)

Before turning to technical issues, we offer a historical perspective. In 1961 John Pierce commented on the *Illiac Suite for String Quartet*, composed by Lejaren Hiller, Leonard Isaacson, and their computer (1957):

> The work of Hiller and Isaacson does demonstrate conclusively that a computer can take over many musical chores which only human beings had been able to do before. A composer. . .might very well rely on a computer for much routine musical drudgery. . .. [T]he computer could be used to try out proposed new rules of composition. . ..
>
> In these days we hear that cybernetics will soon give us machines which learn. . .. Why couldn't they learn what we like, even when we don't know ourselves? Thus, by rewarding or punishing a computer for the success or failure of its efforts, we might so condition the computer that when we pressed a button marked Spanish, classical, rock-and-roll, sweet, etc., it would produce just what we wanted in connection with the terms. (Pierce, 1961 [2nd ed., 1980, p. 260f.]).

Have the last thirty years brought us closer to this vision? The impressive collection on music and connectionism edited by Todd and Loy (1991) and the recent volume by Schwanauer and Levitt (1993) contain ambitious ideas for automatic composition and computer music. The purpose of the present article is more modest: to show how both standard and modern time series techniques can be applied to music. Our focus is more on ideas and methodology than on specific results.

[3] There is some confusion in the literature (and in the editions) about Bach's intended order for the individual pieces of *Die Kunst der Fuge*. We will follow Butler's (1983) numbering scheme, in which the unfinished fugue is Cp. XIV.

[4] The organizers of the Competition selected Cp. XIV as Data Set F and posted it after the official close of the competition in January 1992 because of requests for more data. Since the hope was to inspire creative responses, no specific goal was set for this time series. Although the origin of the "mystery data set" was not revealed until the NATO workshop in May 1992, several participants discovered its source in their explorations. Terry Sanger's prediction went far into the future: he replaced Bach's theme with the theme song from *Gilligan's Island*.

[5] In the late nineteenth century, three separate (and nearly simultaneous) claims were made for the discovery that all three themes of Cp. XIV can be combined with the KdF theme: Higgs (1877), Nottebohm (1881), and Ziehn (1894). See Kolneder (1977), pp. 280ff.

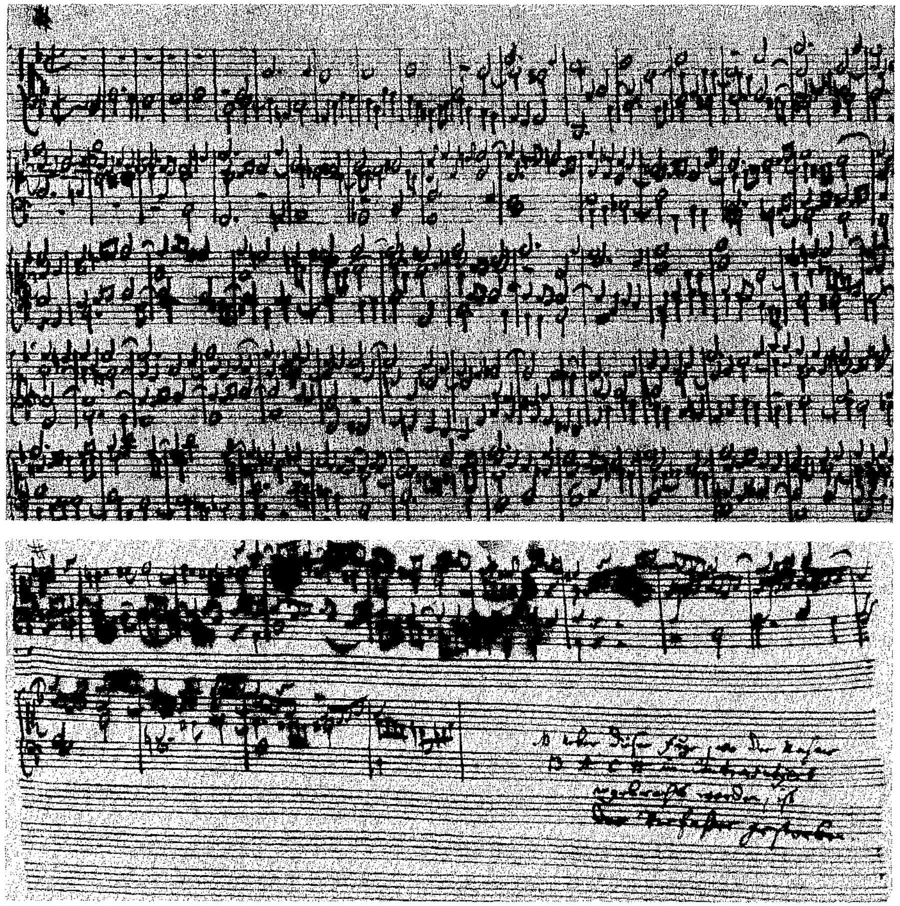

FIGURE 1 First and last page of Bach's manuscript of Cp.XIV. Reprinted with permission.

2. A PHYSICIST'S PERSPECTIVE

What can musicians expect from physicists, data analysts, statisticians, computer scientists, or artificial intelligence workers?

- Computer-assisted *analysis*, ranging from the extraction of the main theme(s) or the structure (e.g., the locations and types of thematic transformations), to the discovery of rules used in composing.

- Computer-assisted *composition*, both for low-level tasks (transitions, episodes, etc.) and high-level structure (such as the possible combinations of themes). In the case of Cp. XIV, we distinguish between two different goals: mere continuation and full-scale completion.

For any of these enterprises, the first step is to choose a representation for the data.

2.1 REPRESENTATION

Musical characteristics vary over seven orders of magnitude in time, extending from 10^{-4} to 10^{3} seconds, as shown in Table 2.[6]

Because of this remarkably large range of time scales, the choice of representation is important. If, for example, we were to choose a representation based on the waveform of a recording, the natural focus of analysis would be on rhythmic or timbral aspects. Since our focus is on the notes themselves, we use three (related) representations that closely resemble the musical text:

TABLE 2 Music spans seven orders of magnitude in time.

characterization	time scales from...	...to
timbre pitch (3 orders of magnitude)	0.0001 sec $(\sim \frac{1}{10\text{kHz}})$	0.1 sec $(\sim \frac{1}{10\text{Hz}})$
rhythm (1 order of magnitude)	0.1 sec (∼sixteenth note in Cp. XIV)	1 sec
melody (1 order of magnitude)	1 sec	10 sec
large-scale form (2 orders of magnitude)	10 sec	1,000 sec (≈15 min)

[6]Seven orders of magnitude in music are large compared to less than half an order of magnitude (one octave) in color perception. However, compared to eleven orders of magnitude present in a river basin with drainage area (Montgomery & Dietrich, 1992), they seem modest.

TABLE 3 The last two measures of `F.dat` in the x-representation. The four voices are indicated by S (Soprano), A (Alto), T (Tenor) and B (Bass). The numbers indicate pitch: a value of 60 corresponds to "middle C." NA denotes a rest. Time t is given in units of sixteenth notes from the beginning of Cp. XIV. The vertical lines (|) indicate bar lines.

```
S |68 68 68 68 NA NA NA NA NA NA NA NA NA NA NA NA|                                               |
A |62 62 61 61 59 59 57 57 56 56 54 54 56 56 59 59|57 57 57 57                                    |
T |54 54 54 54 54 54 53 51 53 53 53 53 53 53 53 53|54 54 56 56 54 54 52 52 51 51 49 49 51 51 54 54|
B |49 49 49 49 49 49 49 49 49 49 49 49 49 49 49 49|42 42 42 42                                    |
   ↑                       ↑                       ↑                       ↑
t  3793                    3801                    3809                    3817
```

- The x-**representation** gives the pitch values for the four voices $(i = 1, \cdots, 4)$ as a function of time. $\mathbf{x}_t$ is the four-dimensional vector at time t, and x_t^i denotes its components. t indicates the time and is given in units of sixteenth notes (semiquavers).[6] Table 3 shows the last two measures of Cp. XIV as it stands in Bach's manuscript.
- For certain tasks, alternative representations can be more appropriate. The **difference representation** is given by $\mathbf{d}_t := \mathbf{x}_t - \mathbf{x}_{t-1}$. The d-series gives the number of semitones of each interval between successive notes. The major advantage of the d-representation over the x-representation is that (exact) transpositions of the theme are identical. The disadvantage of the d-representation is that it ignores absolute pitch—the control that keeps continuations from walking off randomly in harmonic space.
- Alternatively, the **run length representation** gives each note as a pair (p, l)—its pitch number and length. This representation is convenient if rhythm and pitch (or intervals, when the d-series is run length encoded) are to be studied, but it is less suited for analyzing polyphony, since vertical alignment is lost.

All these representations can be augmented with additional explicit information about the placement of each note within the measure. In a connectionist implementation, a sensible metric of similarity is induced if this "phase" is represented by four binary units (the unit corresponding to the most significant bit encodes the location of the note within the first or second half of the measure, etc.).

We first analyze horizontal structure, ignoring the relation of each voice to the others. We then analyze vertical structure, i.e., the interaction of different voices

[6]The fact that this representation does not distinguish between tied and repeated notes is not a serious shortcoming, since the Cp. XIV themes do not contain repeated notes.

at each time step. We then consider polyphonic structure (the unfolding of vertical structure in time) and suggest an automated way of extracting higher level structure. In Section 3, we turn to the more difficult task of continuation.

2.2 HORIZONTAL ANALYSIS (MELODY)

A first step in exploratory data analysis (Tukey, 1977) is to histogram the data. A **histogram** counts the number of occurrences of a specific event.[8] For example, the number of occurrences of each interval in one voice throughout the entire piece can be plotted against interval size (in semitones). Most of the statistics presented in this section can be applied to all three representations given above. However, different representations emphasize different properties, as illustrated by this list of histograms:

1. Distribution of *pitch, allowing for note length*: histogram the x values. This statistic takes the length of each note into acount, e.g., the pitch level of a quarter note is counted four times.
2. Distribution of *intervals* between consecutive notes: histogram d, the number of semitones, from the difference representation. This histogram contains information about relative pitch only.[8]
3. Distribution of *pitch, irrespective of length*: histogram p from the run length encoding. Unlike the first histogram, this method counts the number of occurrences of each pitch, without taking note length into account.
4. Distribution of note *lengths*: histogram l from the run length encoding.

Histograms 1 and 3 can be collapsed over octaves, focusing on pitch class.

In the description of dynamical systems, one-dimensional histograms are used when only minimal information about the system is available. They approximate the probability of states without taking into account any knowledge of the previous state of the system. In literature, histograms have been used in authorship disputes: Thisted and Efron (1987) attribute a poem (discovered in 1985) to Shakespeare by comparing the words in the poem with the entire Shakespeare corpus.

From a horizontal analysis of music, we want more than static statistics. In particular, we want information about temporal progression. The simplest way to collect this information is with a *first-order* **Markov model.**

[8]For a discrete alphabet, the cells of the histogram simply correspond to the characters of the alphabet. Note that there is no "natural" metric between the characters. This is different from histograms of continuous-valued data that are quantized (or binned into the histograms cells).

[8]Fucks (1962) gives the histograms for the x- and the d-representations, as well as the autocorrelation functions. Hsü and Hsü (1990) find that the Fourier transform of the d-histogram obeys a power law (i.e., the log-log plot of the Fourier transform of the intervals resembles a straight line).

Markov (1913) studied the patterns of individual letters in written Russian. He used Alexander Pushkin's *Eugene Onegin* to fill a two-dimensional histogram, along whose sides were the letters of the Cyrillic alphabet. Each time letter i is followed by letter j, the counter in the corresponding (i, j)-cell is incremented. This allowed Markov to extract fuzzy rules for the spelling of Russian.

First-order Markov models, described by two-dimensional histograms or tables, are similar to phase portraits used in the study of dynamical systems. A phase portrait is a plot of x_t against x_{t-1}.

Brown et al. (1992) use 583 million words of written English to build a *second-order* Markov model, which predicts an ASCII character as a function of the two preceding characters. By computing the cross-entropy between their model and a balanced sample of English, they obtain an upper bound for the average amount of information in a printed English character: 1.75 bits. In comparison, the standard Lempel-Ziv algorithm (see Cover and Thomas, 1991) reduces the amount of information from 8 bits per ASCII character to 4.43 bits. The additional compression ratio obtained by Brown et. al. shows that a second-order Markov model, despite its simplicity, captures a significant amount of information about the sequence of letters.

One way to characterize a simple deterministic time series is by the order m of the Markov model, where the next value becomes a single-valued function of the previous m values (i.e., each column in the table of transition probabilities has only one nonzero cell).

Although music is certainly not an entirely deterministic process, histograms can quantify the similarity between pieces, composers, and styles. Suitable histograms can be constructed for pitch (x or p), intervals (d), or length (l). The histograms can be compared in raw form or through summary measures, such as moments, or *entropy* (for one-dimensional histograms), *mutual information* (for two-dimensional histograms), or *redundancy* (for higher orders).[10]

So far, we have dealt with structure that is local in time. A complementary approach is to find structure that is global in time, such as a description obtained by a **Fourier transform,** which yields the spectral coefficients corresponding to the average amount of energy for each segment of the spectrum. Voss and Clarke (1978) take the waveform of a recording (the amplitude as a function of time) and study both its audio power and the rate of zero-crossings ("instantaneous frequency"). They find that the Fourier transforms of both time series are inversely proportional

[10]The redundancy measure describes the information gained by increasing the order of the Markov model. It is based on incremental mutual information as a function of the order of the model (the number of past time steps that are taken into account). Redundancy is a nonlinear generalization of partial autocorrelation, just as mutual information is a generalization of (ordinary auto-) correlation. See Gershenfeld and Weigend (1993).

to the frequency f over several orders of magnitude.[11] An important feature of $1/f$ spectra is that correlations in time are important for the entire range of the spectrum. The Fourier spectra obtained from the audio representation can be compared and contrasted to the spectra obtained from pitch-based representations.

2.3 VERTICAL ANALYSIS (HARMONY)

The previous subsection on horizontal analysis focused on the progression of each individual voice. (The individual voices of a fugue have similar statistics.) We now look at the vertical dimension, ignoring the progression in time. In this section, we use only the x-representation. The term *chord* denotes simultaneously sounding pitches, i.e., the (up to four) pitch numbers in a vertical slice.

The "weakest" model for vertical analysis (the model with the fewest assumptions) simply counts the number of occurrences of each chord. The most simplistic approach is to provide a large array and to increment for each time step the contents of the corresponding cell. The resulting numbers in the array characterize the piece as a whole. They can also be used to assign a "surprise" value to each chord, defined by the negative logarithm of its probability. The number of different chords (i.e., cells that have one or more entries) can be plotted against the total number of chords (i.e., the total number of entries in the array). Gabura (1970) analyzed this average occupation number of the nonempty cells and found the differences between composers to be significant.

Gabura also tried to classify different composers on the basis of pitch structure in their music by using a **neural network** with a binary output unit. His network had no hidden units: backpropagation had not yet been invented. We now show how hidden units allow the extraction of structure from music, and we relate this structure to a number of fields, including cognitive psychology.

An auto-associator neural network is a simple connectionist architecture with hidden units. As a method of encoding the chords, the network is trained to reproduce the input pattern at the output, after piping it through a bottleneck of hidden

[11]Voss and Clarke (1978) analyze the low frequency variations of the audio power by taking the Fourier transform (below 20 Hz) of the squared waveform, after the waveform was bandpassed (100 Hz to 10 kHz). The Fourier coefficient at f (say, 0.1 Hz, to fix an idea) measures the degree of loudness variation at that time scale. The variable in Fourier space, f, is usually called frequency. We have to be careful not to confuse f with pitch: in this context, the Fourier spectrum characterizes temporal periodicities of variations in both loudness and "instantaneous frequency." The spectral information can be presented "back" in the time domain as the autocorrelation function (the inverse Fourier transform of the power spectrum).
A more general, yet still linear technique is **quefrency alanysis** (Bogert, Healy, and Tukey, 1963). It can be applied to each voice in order to extract information such as echo-like repetitions of the themes. It is also interesting to analyze the cross-spectra between voices, i.e., the (complex) covariance of the complex Fourier coefficients between two voices as a function of the relative time delay.

units.[12] Once the network has learned to encode the patterns (i.e., to reproduce them as faithfully as possible, given the limiting number of hidden units), we can try to learn what it has learned. We consider two network responses to a chord at the input: the reconstruction error of the output and the activation values of the hidden units. Standard connectionist analyses of these responses include:

- A scatter plot of *error vs. surprise value* (as defined above). This plot relates the features extracted by the network to the number of occurrences of each chord.
- A plot of the *principal components of the hidden unit activations* (the eigenvalues of the covariance matrix). This plot estimates the effective dimension of pitch space (see Weigend and Rumelhart, 1991).

Since we are dealing with music, we can relate network features—reflecting only the statistics of the musical text—to theories from other fields, including physics, composition theory, and cognitive psychology.

PHYSICS/ACOUSTICS. We live in a world in which most sound generators (vocal chords, string instruments, wind instruments, etc.) are one-dimensional objects, which implies that their spectra contain only integer multiples of the fundamental frequency.[13] To what degree are such physical contingencies reflected in music? Is there structure in a scatter plot of the *network error vs. the spectral overlap* (Kameoka & Kuriyagawa, 1969)?

COMPOSITION THEORY. Eighteenth-century composers followed a general set of rules governing the use of consonances and dissonances (e.g., Fux, 1725). Such "common-practice" rules are implemented in an expert system by Maxwell (1992). He assigns a "dissonance level" to chords (e.g., consonant intervals are assigned dissonance level 1, augmented fifth dissonance level 3, etc.). Is there structure in a scatter plot of the *network error vs. the level of dissonance*? To what degree does the network error "explain" the concept of dissonance, and where are the discrepancies?

[12]There is one input unit for each pitch value in the piece. Each input pattern corresponds to one vertical slice in the x-representation. For each note in the chord, the corresponding input is set to 1; all the other inputs are set to 0. (An alternative representation is suggested by Forte, 1964.) For each chord, the reconstruction error is given by the distance between the target (the given chord) and the prediction (the network output). A general discussion of auto-associator networks can be found in Weigend (1993).

[13]This is different for higher dimensional objects. A circular drum, for example, has noninteger harmonics and subharmonics, located at the zeros of the Bessel functions.

COGNITIVE PSYCHOLOGY. Although pitch numbers are indeed numbers, the proximity of those numbers does not imply perceptual similarity. For example, substituting a C♯ for a C in a C major chord is usually less acceptable than replacing the C with a G, or with a C from another octave. If we want to compute distances between notes with the simple Euclidean metric, the notes must be embedded in a higher dimensional space.

Shepard (1982a,b) constructs a geometrical model that reflects perceived similarity between musical tones. The pitch number x is enhanced by the location on two circles: the circle of chroma describing the sequence $\mathrm{C}, \mathrm{C}\sharp, \mathrm{D}, \ldots, \mathrm{B}, \mathrm{C}$, and the circle of fifths $\mathrm{C}, \mathrm{G}, \mathrm{D}, \ldots, \mathrm{F}, \mathrm{C}$. Pitch and one of these two "wrap-around" variables can be visualized as a helix in three-dimensional space. By adding the other cyclical variable, we can construct a helix of a helix in five-dimensional Euclidean space. Although all five coordinate values are functions of only one pitch number, this embedding allows the Euclidean norm to do justice to cognitive-structural constraints.

Shepard shows that listeners' judgments induce a metric in pitch space. The activation values of the auto-associator's hidden units also induce a metric. Are the *similarities in network response* related to Shepard's?[14]

The network encoding/decoding error can also be used to characterize the temporal evolution of the fugue. We plot a smoothed version of the *error vs. time*. (Short-term fluctuations are removed by averaging over a bar of music or by applying a standard smoothing convolution filter or some denoising by wavelets.) Furthermore, we plot the *volatility* of the error as a function of time. (The volatility is the running standard deviation of the errors computed in a sliding window in time; see Weigend et al., 1992, p. 419.) These ideas can be traced back to Jackson (1970), who plots a "rate of dissonance" against the measure number.

2.4 POLYPHONIC ANALYSIS

We now come to the most important part of fugue analysis: polyphonic structure—the unfolding in time of vertical elements (i.e., how the individual voices fit together and relate to one another). Music theorists of the Renaissance and Baroque wrote innumerable treatises on counterpoint, each elaborating the various techniques and rules that govern the combination of two or more melodic lines (the essence of counterpoint). Many of these rules are formulated as prohibitions: "Do not use parallel fifths or octaves." This approach can be contrasted to a data-driven analysis, which finds (often implicit) generative rather than restrictive rules.[15]

[14] Useful methods that can be used to relate these two representations include clustering and multidimensional scaling, as well as visualization with the help of self-organizing feature maps (Kohonen, 1990).

[15] We are aware of only one computer program for automatic counterpoint (Schottstaedt, 1989).

The following three examples of four-part chorale harmonization in the style of J. S. Bach illustrate the transition from artificial intelligence (AI) without learning, through traditional AI with learning, to connectionism. Ebcioglu (1988) uses an expert system with some 350 rules for a "generate-and-test method." These rules were hand-coded in a form of first-order predicate calculus—not learned. Schwanauer (1993) describes a rule-based system with a chunker that combines successful sequences of rules to new rules. Hild et al. (1992) take a connectionist approach. Their neural network has 70 hidden units. By learning to extract regularities from the training examples, it produces convincing harmonizations for new melodies.

The task of chorale harmonization differs from the goal of automatic continuation in one important respect: in chorale harmonization the melody is always given, whereas in automatic continuation there are no pre-existent parts. The key to automatic analysis and continuation is the recognition of not only small-scale patterns, but also higher order structure. In the next section, we suggest a method of extracting higher order structure.

2.5 HIGHER ORDER STRUCTURE

In a traditional fugue analysis, one of the first steps is to look at thematic structure—that is, to identify the themes, their recurrences, and transformations. Let us pretend we do not know the themes: can we extract them from the data alone? We suggest an automated analysis based on **clustering**:

1. Use the difference representation. Define a window of length w. (For example, in the Cp. XIV data set, setting w to 32 or 48 corresponds to two or three measures, respectively.) Start from the first note of the first voice. Each time the window is "sat down" on the data, it produces a point in w-dimensional space. Record the first point. Advance the window by half a bar or a full bar (8 or 16 time steps), generating the next point in w-dimensional space, and continue through the entire piece with one voice after another. This will produce several hundred points in the w-dimensional space.
2. Cluster these points. The clusters with low variance and a relatively large number of points correspond to parts of the fugue themes. (There is some variation within these clusters because tonal answers require slight modifications.)
3. Treat each cluster center as a symbol. Since each point in the w-dimensional space is assigned to one of the clusters, the original series is now transformed into a sequence of symbols. Although the number of significant clusters is comparable to the number of pitch values, this procedure captures structure at a time scale an order of magnitude slower than the original representation, because the window is moved by 8 or 16 steps every time.
4. Analyze the transitions between these symbols. Some of the symbols are followed consistently by a single symbol. These pairs correspond to adjacent parts of one of the themes. They can be combined into *compound symbols* (Simon &

Sumner, 1968; Redlich, 1993). Symbols which have a large number of successors (stochastic transitions) signal the end or the absence of a theme.

In order to visualize the results of this automatic analysis, we can construct matched filters (corresponding to the cluster centers), convolve them with the data, and plot the results. If we are lucky, the thematic structure of the fugue will appear.

3. CONTINUATION AND EXPECTATIONS

Problems in scientific analysis and inference have two dimensions: theory and data. Traditionally, music analysis has been theory-rich and data-poor; music theorists seldom construct their theories from the data ("bottom up"). Because of the widening availability of musical texts in computer readable form,[16] musical analysis can now incorporate data-rich modeling. Data-rich/theory-poor modeling starts from the data, not from first principles or theories. We use the data to construct a model that makes predictions. We then analyze where the predictions went wrong, modify the model, predict, analyze, modify, etc. This method is sometimes called **analysis by synthesis.**

In the case of Cp. XIV, the idea is to build a model from Bach's fragment, generate continuations, and analyze the shortcomings. This approach will suggest improvements for the model. It may also broaden our understanding of human cognition and musical creativity. In Section 3.1 we list some approaches to continuation; in Section 3.2 we address the question of how expectations, central in music cognition, might be modeled from the data. In Section 4 we contrast this inductive approach to the deductive, approach of traditional musicology.

3.1 CONTINUATION

A straightforward implementation of an inductive approach is to find a part of the past that resembles the present, and to predict the same continuation. This **nearest-neighbor** approach taken by Zhang and Hutchinson (this volume).[17]

[16]The complete works of J. S. Bach (and works by other composers) will soon be available from the *Center for Computer-Assisted Research in the Humanities* (Hewlett and Selfridge-Field, 1989; Hewlett, in preparation).

[17]Zhang and Hutchinson (this volume) use run length encoding. They consider the last two notes in all four voices (pitch values and length $\times$ 2 previous values $\times$ 4 voices = 16 numbers). In this 16-dimensional space they find the 10 nearest neighbors in the training set to the present point. For a continuation (of 4 + 4 dimensions, i.e., the next values for pitch and length for the four voices), there are two possibilities: (1) to stochastically pick one of the past continuations (with equal probability or with a probability that reflects the distance), or (2) to use an average of the ten neighbors, rounded to integers. One problem is the choice of metric in this space of pitch $\times$ length. Zhang and Hutchinson use the sum of the component-wise differences. Unfortunately, this mixes pitch with duration. Even in pitch space by itself, this metric is inappropriate. Better representations are discussed in the main text. Furthermore, run length encoding destroys any vertical structure.

Another approach is to express the next value in the series as a function of previous values. This is called an **autoregressive model** (AR model). Assumptions have to be made about the interpolating function: in the simplest case, it is linear.[18]

Feedforward **neural networks** with sigmoid hidden units are a nonlinear generalization of linear AR models. The input into each hidden unit can be viewed as a linear filter, the activation of the hidden unit as a squashed version of the filter, and the output of the entire network (its prediction) as a weighted superposition of the squashed filters. The filter coefficients and the weights to the output are adjusted to fit the training data.

In the last four years, connectionist networks have successfully emulated dynamical systems and predicted their time series. The application of connectionist networks to music, however, differs in two important respects:

1. In dynamical systems, the squared difference between the predicted and the target value is a reasonable error measure. Music is different: a semitone error is usually worse than an octave error.
2. For systems governed by differential equations, a sufficiently large number of past values provides all information necessary for prediction. Music is different: it has structure on a hierarchy of time scales.

The first point can be addressed in two ways: by using an appropriate metric that reflects cognitive-structural constraints, or by avoiding a metric altogether. Shepard's representation (see Section 2.3) can serve as a "good" metric; it has been used by Mozer (1991) in a network to generate Bach-like tunes. There are two ways of avoiding a metric: Markov models and a local representation.

Markov models need not assume any metric or distance function between the symbols.[19] Having weak assumptions requires large amounts of training data: the training set size increases exponentially with the order of the model ("curse of dimensionality"). In order to learn from data without having an intractable number of cells to fill, Kohonen et al. (1991) use a *dynamically expanding context*. The size of the context starts at zero and is expanded until either all ambiguity has been

[18] Musha and Goto (1989) use a linear autoregressive model for Schubert's *Gute Nacht*. Their filter takes the past 64 pitch values in an x-representation into account; each time step corresponds to an eighth note. Creating different pieces with the same filter is equivalent to using surrogate data with a spectrum smoothed by a fit with 64 parameters. In surrogate data, the idea is to Fourier transform a time series into frequency space, to randomize the phases (keeping the amplitudes), and to inverse Fourier transform back to the time domain. The power spectrum (the squares of the amplitudes) of the surrogate series is, by definition, identical to that of the original series. See Theiler et al. (this volume) and Kaplan (this volume).

[19]A review of Markov models for composition (from a nonlearning perspective) is given by Ames (1989).

resolved (deterministic continuation) or until a maximal context length (8 in their run length encoding) is reached—whichever comes first. Todd (1988, 1989), who also tries to avoid an inappropriate metric, uses a *local representation*: each unit corresponds to a single pitch value, similar to the representation discussed above.[20]

We turn to the second point—the hierarchy of time scales. If we want to use an ordinary Markov model or a feedforward network, we can explicitly incorporate knowledge about the hierarchical structure (obtained, for example, from the clustering method discussed in Section 2.4).[21] **Hidden Markov models** and **recurrent networks** have more computational power than ordinary Markov models and feedforward networks because the former can learn to represent the past internally. Mozer (1993) gives an overview of different recurrent network architectures and discusses their advantages and disadvantages.

3.2 EXPECTATIONS

In the Introduction we mentioned the importance of the tension between randomness and order. For music to "work," some balance has to be struck between the realization and the violation of deterministic predictions (expectations).

Meyer (1956) is the foremost exponent of the metaphor of expectations (or expectancies) in music criticism. His method of musical analysis is based on music's tendency to arouse expectations on both large and small time scales. Unfortunately, his ideas remain peripheral to most music theorists and critics: it may be that expectations are too subjective or too imprecise for traditional music analysis.

Nevertheless, the fact is that musical expectations do exist. However, the question of how they arise—as a function of both past musical experience and present input—is not easily answered. Expectations generated by a musical phrase may be due to a number of historical factors, ranging from common musical practices to peculiarities of specific traditions, schools, or composers. Other influences on musical expectations include the myriad musics (and musaks) of modernity.

Leonard Bernstein, in a discussion of Beethoven's Sixth Symphony, articulates the "formalist" method of looking at musical expectations:

> We are concerned not with the birds and bees, but with the F's and G's, the notes themselves which form the intrinsic metaphors of music, metaphors that evolve out of syntactic and phonological transformations. (Bernstein, 1976, p. 154)

[20]Todd's goal is to generate tunes. His network differs from the chord encoder discussed in Section 2.3. in three ways: his network predicts the next note (rather than the same chord), it learns to represent the past internally through recurrent connections, and it has some extra inputs that represent musical style.

[21]Cope (1991) approaches this problem with a hybrid system that does not fit our classification. His system composes "recombinant music" by chopping up a piece of music and recombining the parts in new ways.

By relating music to Chomsky's ideas on linguistics, Bernstein suggests that music ("the F's and G's") is more crucial than an extraneous program ("the birds and bees"—the representational reading of Beethoven's *Pastoral Symphony*). Like Bernstein, we are more interested in the notes and their resulting structures than in extra-musical information. But unlike Bernstein, our interest in the notes is purely statistical. We want to see whether musical expectations can be derived solely from the musical text.

There have been some recent experiments that model the formation of expectations by using synthetic (computer-generated) data. In music, Bharucha and Todd (1989) model tonal expectancy with connectionist techniques. They use series of isolated chords in succession, and sequences of seven successive chords each.

In linguistics, Elman (1990) generates sentences with a simple grammar, using a set of one thousand words. His network predicts the next letter, differing from the auto-associator presented above in Section 2.3 in two respects: Elman's network is trained to predict the next step ("hetero-associator"), and it contains recurrent connections that encode relevant features of the past. The network generates an expectation that reflects the probability of the next letter. An error is obtained by comparing the prediction with the actual letter. A large error indicates a violation of the expectation and often signals a boundary between words.

With a large amount of music available in computer-readable form, real data (rather than computer-generated data) can be used to build models for expectations in music. A data-driven approach may help separate the truly creative from the merely mechanical, and thus distinguish "Bachian Creativity" from "Bachian Noise."

We have not yet addressed the issue of how to complete Bach's unfinished fugue. But a fugue has a beginning and an end—unlike dynamical systems theory, where initial transients are usually considered to have decayed and the system is assumed to be in a stationary state. If the computer is to complete Bach's last fugue, it needs to know how Bach completed other fugues. With all of Bach's fugues (and more) in the computer, will modern learning algorithms on powerful machines be able to generate satisfactory completions for Cp. XIV?

The challenge posed by this question is hardly new; many musicians have struggled with Bach's most ambitious fugue and admitted defeat. Others have chosen to play the Baroque forecasting game, and have left their mark upon Cp. XIV with published completions and elaborate detective stories.

4. A MUSICOLOGIST'S PERSPECTIVE

4.1 PUBLISHED COMPLETIONS TO CP. XIV

Composers, musicologists, and musicians of all stripes have taken turns completing Cp. XIV: there are numerous published endings by an array of Bach "wannabes." Most of the completions share a common thematic content—the three fugue themes from Cp. XIV plus the KdF theme. There is no such agreement on the scope of the missing portion, although Butler (1983) has determined how many pages were allotted for Cp. XIV in the plan for the original edition.[22]

The issue of overall length must be decided before completion can be attempted. Cp. XIV, as it stands in Bach's manuscript, is a fugue in three large sections, each of which has a different theme. The manuscript breaks off in measure 239, on the heels of the first combination of all three themes, a maneuver which may be the second part of section three (the B.A.C.H. section)—or, the beginning of an entirely new closing section.

Some latter-day Bachs bring in the KdF theme right away, making the third section the final section. Others extend the third section beyond measure 239 (where Bach stopped) and include an entirely new closing section in which all four themes are combined. The partisans of the three-section approach include a number of first-rate scholars and musicians: Tovey (1931), Walcha (1967), Wolff (1975), Butler (1983), Moroney (1989), and Schulenberg (1992). Wolff and Butler present solid arguments for a fairly brief concluding section, while Tovey, Walcha, Moroney, and Schulenberg each try their hand at composition. The four-section sympathizers—Busoni (1912), Husmann (1938), and Bergel (1985)—are less numerous, but considerably more adventurous. Using Bach's other multisection fugues as models,[23] they prefer to bring the third section to a close with a full cadence before bringing in the KdF theme in a new fourth section.

The various completions range from the perfunctory to the outrageous: Schulenberg provides a four-bar solution (for the faint of heart), while Bergel proposes a monumental four-section fugue of some 381 measures. There are, predictably, a few iconoclasts in the crowd: Busoni embeds Bach's fugue within a gigantic improvisatory fantasy for piano (44 pages worth!), while Martin (1948) offers two different completions, each of which merely completes section three without introducing the KdF theme.

[22]This is not to say that Cp. XIV was conceived to fit into the allotted space (six pages by Butler's calculation) but rather, that six pages were allotted for it. This fugue was probably not complete at the time of the original pagination scheme, and the six-page requirement may have been one of the reasons why Bach never finished (or gave up copying) the piece. (This issue is discussed later in the main text.)

[23]Among others, Cp. VIII and XI from the KdF and the E♭ organ fugue, BWV 552b.

4.2 THE MUSICOLOGICAL "DETECTIVE STORY" OF CP. XIV

Most of the published completions to Cp. XIV follow traditional methods of composition.[24] But musicologists are seldom composers; they tend toward primary source study. This type of inquiry has created a never-ending, ever-changing detective story that seeks to "explain" Bach's unfinished fugue. In keeping with this tradition, we suggest yet another answer to the question, "Why did Bach stop?" We return to Bach's manuscript of the unfinished fugue with the following questions in mind:

1. What kind of a score is Bach's manuscript of Cp. XIV? (Is it a composing score, a printer's final copy, or something in between?)
2. Why is Cp. XIV in two-stave format? (All the other contrapuncti are in four-stave format.)
3. How can we explain the infamous final bar? (Why did Bach stop writing in the middle of the fifth page without any remark?)

The first question invites a comparison with another movement from KdF. The final version of the *Canone per augmentationem in motu contrario* is the only other piece in Bach's personal copy of KdF written on the same type of paper as Cp. XIV. Both pieces are on loose sheets; they were either revisions or additions to the collection of stitched folios known as the *P200* manuscript. The two separate movements are on oblong paper with five systems of two staves each. Bach had a good reason for making a second version of the canon: the final version is a rearrangement (in larger note values and in ₵ time) of the earlier *Canone per augment. in motu contr.* in P200 (which was in C time).

Bach's ₵ time version of the canon is clearly the final version: the cleanness of the copy and its layout—a page turn falls conveniently when one hand is resting—are proof that this was the version Bach wished to print. Since this loose-leaf copy of the canon is a revision, we might assume that the existing fragment of Cp. XIV is also a revision—they are both laid out in the same two-stave format on the same type of paper, after all. But if the Cp. XIV manuscript is a revision copy, what did Bach revise?

A change in meter (the reason for the revision of the canon) does not seem to be the right answer. A reduction of note values in the first section of Cp. XIV would put the fugue subject and the answer into different metric positions—not a likely possibility. Nor will the first exposition of Cp. XIV work in ₵ time (with two whole notes per bar, as in Cp. I, II, and III); there is the same problem of shifting metric position for each entry of the theme.

There is, however, one clue in the manuscript of Cp. XIV that supports the hypothesis that there was some sort of metric change between versions. On the last two systems of the fourth page (in the B.A.C.H. section), Bach writes partial

[24]Weigend and Dirst (in preparation) address the question of whether the computer can statistically "authenticate" the completions listed in the Appendix.

bar lines every two bars: precisely the sort of error one would expect if Bach was copying from an earlier version in a different meter—particularly if that meter was ₵ time with four half notes per bar. In the midst of copying the B.A.C.H. portion into his manuscript, Bach must have momentarily forgotten that he was (now) in ₵ time; he inadvertently reproduced the barring from the previous version of the (lost?) B.A.C.H. fugue.[25]

How can we explain Bach's mistake? Perhaps the B.A.C.H. section was written first. This would fit in nicely with Wolff's idea concerning "Fragment X" (the lost completion), which Wolff supposes Bach must have written (or at least sketched out) first, *before* composing sections one through three. Perhaps Bach's procedure for this fugue was totally backwards: he may have worked out the concluding section first (as Wolff suggests), and only then figured out how he was going to get there. If B.A.C.H. had to appear in third place, why not use an already existing fugue on this theme?

Our answer to the first question—What kind of score is this?—thus posits the following scenario: Bach began work on Cp. XIV by inventing three new themes that could be combined with the KdF theme. He then resurrected an earlier fugue exposition on B.A.C.H.[26] Finally, after planning the general shape of the entire fugue, he composed the rest. Bach's manuscript of Cp. XIV is, then, a combination of revision and composing score.

Our answer to the second question—Why the two-stave format?—begins with Butler's conclusion that Cp. XIV was supposed to fill six pages in the original publication. Butler's reconstruction of the (lost) original pagination scheme is quite clever; it neatly solves some lingering problems associated with the whole work.[27] But Butler does not answer all the questions. If Cp. XIV was supposed to serve as the final contrapunctus (Butler's idea, after all), then why is it not in four-stave format like all the other contrapuncti? Even if Bach was revising, why would he have condensed the score at the same time?

[25]There are no other fully authenticated examples of B.A.C.H. fugues by J. S. Bach: BWV 898 is spurious; BWV Anh. 45 has been ascribed to both Justin Heinrich Knecht (Schmieder, 1990, p. 899) and Johann Christian Bach (Kobayashi, 1973, p. 391); BWV Anh. 107, 108, and 110 are probably the work of Georg Andreas Sorge (Schmieder, 1990, p. 920); BWV Anh. 109, the only other B.A.C.H. fugue in Schmieder, is stylistically so anomalous as to be irrelevant for serious comparison with the B.A.C.H. portion of Cp. XIV. Although Bach may have never written a complete fugue on B.A.C.H., he used this chromatic motive occasionally in larger works. For a listing of B.A.C.H. motives in works of J. S. Bach, see the preface to *B.A.C.H. Fugen der Familie Bach*, edited by Fedtke (1984).

[26]Schulenberg (1992, p. 368) also wonders whether the various sections of this fugue were composed separately. He ventures that "Bach composed [Cp. XIV] in sections, linking them by bridges that perhaps were worked out only during the writing of the surviving autograph."

[27]Bach died before the first printed edition of the work was finished, and the executors of his estate (C. P. E. Bach and Agricola) misunderstood his intentions for the projected KdF publication. The confused state of the original (1751–2) publication explains the radical differences between (even recent) editions of KdF.

The most likely reason is that Bach was behind schedule and needed to save some space. We have no way of knowing whether Bach finished Cp. XIV before the pagination scheme was drawn up. But if we suppose that he had not, the two-stave format can be easily explained. Bach's printer, having already engraved Cp. I–X, needed to know the order of the remaining pieces in the volume. Cp. XIV, for which Bach or his printer (or both) allotted six pages, was to appear after Cp. XIII. When Bach finally began composing (or perhaps just revising and copying) Cp. XIV, he knew he had only six pages for the job. This may have been why he changed the format: in two staves, the work occupies less space.

The two-stave arrangement evidently worked fine until page 5, where Bach ran out of properly ruled two-stave paper (in the middle of the B.A.C.H. section) and was forced to continue on a badly ruled, smaller piece of paper. (Page 5 of the manuscript, reprinted in Figure 1, is about 1 centimeter less in width and length than the previous four pages.) Midway through the second system on this page, after the (first and only) combination of all three themes, the music stops. The appearance of the final measure is puzzling: Bach wrote the tenor part through to the next bar line, while the other voices simply stop on the downbeat of measure 239.

Without a *nota bene* indication or Wolff's "Fragment X" (which would have to begin with the missing voices in measure 239), we may never be sure what Bach's intentions were. But we can make an educated guess: Bach was copying from an earlier version of this fugue (or at least a sketch of the fugue's completion), and he probably did his copying much like we do—one voice at a time, bar by bar. The last measure (the subject of our final question) thus can be understood if we realize that Bach was *copying, not composing*. In measure 239 Bach copied only one voice and stopped (for whatever reason) without copying the other parts through to the end of the measure.

Now for the big one: Why did he stop? Perhaps he was unhappy with the lousy paper (Bergel). Perhaps he did not need to recopy a completion that already existed (Wolff). Or perhaps Bach realized that to continue would be foolhardy, because the fugue was already too long and would never fit within the allocated number of pages. He copied as far as his previous copy went and was unable (or unwilling) to complete this tour-de-force fugue, choosing instead to exclude the five separate sheets of Cp. XIV from the P200 manuscript. Who knows: he may have deliberately destroyed Wolff's "Fragment X" in order to keep anyone from trying to publish the unfinished final contrapunctus of his *Kunst der Fuge*.

Bach's failure to complete Cp. XIV need not doom our efforts; it may, in fact, be an instructive lesson in humility. Although we have no solution yet, the desire to complete this (in)famous work remains strong, particularly since its completion can now be considered from more than just a purely musical perspective.

ACKNOWLEDGMENTS

We thank our friends and colleagues for all the discussions related to this enterprise, in particular Neil Gershenfeld (MIT Media Lab) and Eckhard Kahle (IRCAM).

APPENDIX

Contrapunctus XIV and some completions are available at `ftp.santafe.edu` in the directory `pub/Time-Series/Bach`. Table 4 lists the file names, composers, year, and number of measures for each completion. Bach's unfinished measure 239 is counted as measure 1 of each completion. We invite further "predictions." Please contact one of the authors with your continuations, suggestions, questions, or comments. General directions for accessing the data of the Competition are given in the Appendix of the volume.

Note. A version of this paper has been submitted to *Music Theory Online* (MTO) at `mto-serv@husc.harvard.edu` .

TABLE 4 Cp. XIV in different representations and some completions.

file name	composer	year	length (in measures)	comments
README				description of files
F.dat	J. S. Bach	c. 1750	239	x-representation
F.dif				difference series
F.rl1				run length (soprano)
F.rl2				run length (alto)
F.rl3				run length (tenor)
F.rl4				run length (bass)
Tovey	D. F. Tovey	1931	79	
Martin1	B. Martin	1948	52	
Martin2	B. Martin	1948	41	
Walcha	H. Walcha	1967	72	
Bergel	E. Bergel	1985	143	
Moroney	D. Moroney	1989	31	
Schulenbg	D. Schulenberg	1992	42	

II. Time Series Prediction