Comparing human and computational models of music prediction

by Ian H. Witten, Leonard C. Manzara & Darrell Conklin

Comparing Human and Computational
Models of Music Prediction

Ian H. Witten

Department of Computer Science, University of Waikato, Hamilton, New Zealand

Leonard C. Manzara

Department of Computer Science, University of Calgary, Calgary, Canada

Darrell Conklin

Department of Computer Science, Queens University, Kingston, Canada

Abstract

The information content of each successive note in a piece of music is not an intrinsic musical property but depends on the listener's own model of a genre of music. Human listeners' models can be elicited by having them guess successive notes and assign probabilities to their guesses by gambling. Computational models can be constructed by developing a structural framework for prediction, and "training" the system by having it assimilate a corpus of sample compositions and adjust its internal probability estimates accordingly. These two modeling techniques turn out to yield remarkably similar values for the information content, or "entropy," of the Bach chorale melodies.

While previous research has concentrated on the overall information content of whole pieces of music, the present study evaluates and compares the two kinds of model in fine detail. Their predictions for two particular chorale melodies are analyzed on a note-by-note basis, and the smoothed information profiles of the chorales are examined and compared. Apart from the intrinsic interest of comparing human with computational models of music, several conclusions are drawn for the improvement of computational models.

Please address correspondence to

Ian H. Witten
Department of Computer Science, University of Waikato, Hamilton, New Zealand
Email ihw@waikato.ac.NZ; phone (+64 7) 838–4246; fax (+64 7) 838–4155

May 3, 1992

1 Introduction

The information content of each successive note in a piece of music is not an intrinsic musical property but depends on the listener's own model of the genre of music to which the work belongs. We quantify the information content in terms of "entropy," a measure of the disorder or unpredictability of a system. When applied to any communication situation, entropy measures the amount of information contained in a message. As messages become more disordered, that is, less predictable, it becomes more difficult to transmit them efficiently. Disordered messages require more symbols in their representation, while ordered ones can be coded with fewer symbols.

Perceptive musicians have an instinctive feel for the amount of information communicated in the music they play and hear. For example, most would agree that the music of Schoenberg or Carter is much more information-laden than that of Glass or Haydn. Within a composer's oeuvre, similar distinctions can be made. The late quartets of Beethoven, for example, are considerably denser, in the informational sense, than the early quartets. Meaningful stylistic comparisons between individual pieces, between composers, and between genres can be undertaken by objectively measuring perceptual complexity in terms of entropy.

At a finer level of detail, the amount of information conveyed in a particular piece of music varies as the piece progresses. This is instinctively felt, for example, in the development section of a sonata, or in the stretto section of a fugue. Descending further, the information content varies markedly on a note-by-note basis, some notes being quite predictable while others are quite unpredictable—that is, "surprising." This variation is reflected in the "entropy profile" of the work, which measures the information flow as the piece progresses. Entropy profiles give the music theorist a valuable new tool with which to characterize individual works.

This paper assesses and compares the entropy of musical style with respect to two different models of music: human and computational. The former can be elicited by having people guess successive notes in a piece and assign probabilities to them by gambling [4]. The latter can be constructed by developing a structural framework for prediction, and "training" a modeler by having it assimilate a corpus of music and adjust its internal probability estimates accordingly [2]. In this latter case, we can examine in detail how the model views particular musical events by inspecting its inner workings at each stage of computation. With a human model, direct examination of the mind's expectations is not possible, and we must adopt a more indirect strategy to measure the

entropy with respect to the model.

For the purpose of musical analysis it would be attractive to distinguish *stylistic* entropy, or the uncertainty inherent in the musical style, from *perceptual* entropy, or that which is relative to the listener's model and fluctuates from one listener to another. One might further want to distinguish *designed* entropy in order to account for the composer's intentional deviations from the stylistic norm. Unfortunately, it is not possible to distinguish between these measures in practice—and it is a matter of some debate whether the distinctions hold even in principle. The present work measures entropy only with respect to a model. In the case of human listeners, this is their perceptual model. The fact that there is no single human model complicates the task, and in the experiments outlined below we increase the reliability of the entropy estimate by basing it on the performance of several human subjects.

The two modeling techniques have been described and evaluated separately on the Bach chorales [2] [4], and turn out to yield surprisingly similar values for the entropy of their melodies. Previous research has focussed on the overall information content of whole pieces of music. In contrast, the present study evaluates and compares the two kinds of model in fine detail in terms of the individual predictions they make.

The next section gives a very brief introduction to the idea of modeling and prediction. Following that we review human and computational models of music prediction—both the methodology of eliciting subjective probabilities of predictions from people, and the construction of a computational model that accommodates different musical dimensions, or perspectives, and can be trained by exposure to particular pieces of music. Section 4 comprises the main body of the paper: an analysis of the predictions of these models on two chorale melodies. We present and compare the predictions on a note-by-note basis, from a musical point of view, and, in addition, we examine and compare the overall information profiles of the chorales according to the models. The next section teases out some implications for the computational model, including several suggestions for its improvement.

2 Models of Prediction

The entropy of any sequence of events is measured with respect to a "model" of potential continuations of the sequence. Models can take many forms, including ones that develop naturally

in our minds over many years of training, formal models that are synthetically constructed from a theory of music, and computational models that use adaptive techniques of learning. All share the ability to predict upcoming events based on what has happened previously.

Models predict continuations by assigning a probability to each possible next event. For example, if the events x, y, and z comprise the possible continuations, a model might say that event x is 50% likely, event y 10% likely, and event z 40% likely. Such a model is never categorically "wrong" unless it assigns a probability of 0% to an event that actually occurs.

An entropy profile for a sequence of events can be derived by examining the probabilities that the model assigns to the events that actually occur in the sequence. If an event is given a high probability, it contributes only a little to the total entropy of the sequence. If it is given a low probability, then it contributes a great deal. The profile thus consists of an event-by-event plot of entropy against time.

Entropy is usually expressed in terms of bits per event. A bit is a binary digit, with either the value 0 or 1. If an event has an entropy of 1 bit, this implies that the model assigns a probability of 50% to that event. If its entropy is 2 bits, then the model assigns it a probability of 25%. In general, the probability p of an event with entropy e bits can be found from the formula $p = 1/2^e$ (or, equivalently, $e = -\log_2 p$).

Given different models of music, it is natural to ask "Which model is better?", or, more operationally, "Which performs the best?" Such a determination can be made by examining the average entropy produced by the model as it predicts each event of the sequence. The lower the value, the more successfully the model has "understood" the sequence of events under scrutiny. Conversely, the higher the average entropy, the less successful the model is at prediction. We thus equate the quality of the model with its predictive ability measured in terms of entropy.

By comparing entropy profiles, one can spot particular weaknesses of one model with respect to another. If it has difficulty predicting certain events in a sequence (indicated by peaks in its entropy profile that are not present in the other model's profile) then the model can perhaps be modified to better handle such events. For a computational system, this may mean altering the particular algorithms it uses. In the case of humans, the model might be improved by providing instruction.

3 Human and Computational Music Prediction

We became interested in human models of music prediction through our research into the creation of a powerful, machine-based predictive model of music using machine learning [2]. There were two reasons why we needed more information about human performance. First was the compelling question of determining just how good the model was. Although its performance is easily measured in the abstract, these measurements hold little meaning if we do not know how well people perform at the same task. The second reason is more technical. As described below (Section 3.2), the system uses several musical dimensions, called "viewpoints," to guide its adaptive learning abilities. A major challenge has been to decide which viewpoints are most important in developing an effective model, and in what manner they should be linked. At first we thought that these choices could be guided by standard music theory alone, but it soon became clear that a precise, more objective, method was needed to tune the model. These two questions prompted the experiments described below.

3.1 Eliciting Predictions from People

Around 1950, Shannon, the father of information theory, sought entropy estimates for ordinary people's models of plain written English text [5]. He did this by eliciting predictions from people, in the form of a guessing game. The procedure was:

1. Show a passage of text to a subject, but only up to a certain point;

2. Ask the subject to guess the next letter;

3. If the guess is wrong, have the subject guess again;

4. Repeat until the guess is correct;

5. Have the subject guess the following letter in the same manner, continuing until the entire passage has been completed.

The number of guesses taken by the subject for each letter of the text is recorded. From this information an estimate of the entropy of English with respect to the subject's model can be derived.

4

The drawback of this technique is that it can only yield upper and lower bounds to the entropy estimate, and these turn out to be quite far apart. The guessing methodology was refined by Cover & King [3] by adding "gambling" in order to elicit subjective probabilities for the next letter. At each guess the subject is required to bet a proportion of their capital on the choice. If the prediction turns out to be correct, they are paid fair odds. If it is incorrect, then their capital decreases by the amount bet. The total capital amount accrued by the subject is used to derive an entropy estimate. The results from several subjects can be combined using a "committee gambling estimate," which is often lower than the best individual subject's estimate.

We adapted this gambling methodology to derive an entropy estimate of the pitch of selected Bach chorale melodies, as described more fully in [4]. A computer program called CHORALE CASINO was written to facilitate these experiments. The program is fully interactive, and automates the play of the game by displaying the music, recording bets, and calculating winnings. Each player's actions are fully logged, and from this record we can derive entropy profiles for each chorale examined. The profiles for Chorales 151 and 61 are analyzed below.

3.2 Multiple-Viewpoint Model of Music

Can a machine generate acceptable specimens of a musical style? Our research on computational models of music has taken two rather unusual approaches to the question [2]. First, we use *predictive power* as an objective yardstick of generative ability. This is measured by the compression that our system gives to a set of test cases, and is determined by calculating the average number of bits needed to represent them. Second, rather than basing the model on rules and constraints that are elicited from experts or musical treatises on the style, we use machine learning techniques to automate its construction. A large corpus of training cases is shown to the machine, which incrementally specializes its model by incorporating them into a knowledge base.

The main features of any computational model of music are a representation for musical knowledge and inference methods that operate on that representation. We choose to view the musical surface from many different perspectives or *viewpoints*—for example, as sequences of melodic intervals, pitch classes, or contour indicators. Each viewpoint is an independent knowledge source that makes inferences about some feature of the musical surface.

The multiple-viewpoint scheme is defined formally in [2]; a brief informal description follows.

The musical surface of a chorale is represented as a sequence of discrete events. Each has a start-time, a duration, a pitch, and a fermata indicator. These constituents are called *basic types*. A viewpoint derives new types from sequences of events, and views the musical surface as a string of these *derived types*.

For example, a *contour* viewpoint regards the surface as a sequence of elements in the set $\{-1, 0, 1\}$ according to whether the pitch falls, remains unchanged, or rises, from one note to the next. The elements of an *interval from referent* viewpoint are descriptors in the set $\{0, \ldots, 12\}$, giving the interval in semitones of the note above a reference pitch. Viewpoints can be linked together to form new ones. For example, a viewpoint linking the *contour* and *interval from referent* viewpoints has elements in the cartesian product set $\{-1, 0, 1\} \times \{0, \ldots, 12\}$. Linked viewpoints model correlations between descriptors.

A machine learning system for music, including about twenty viewpoints, has been implemented in Common LISP. During the learning phase each viewpoint divides the music into a set of fixed length *traces* and incorporates these into the knowledge base. A trace is simply a sequence of contiguous viewpoint elements. The knowledge base is initially empty: each incorporation specializes it slightly to the musical style. Statistics are kept on all traces and their subtraces, and they are indexed for efficient retrieval in the future. The incorporation algorithm is designed to avoid overspecializing the theory to the examples. After processing a large sample of traces, each viewpoint will have developed a constrained probabilistic model for its musical perspective.

A viewpoint makes inferences about an incomplete piece by taking as context the events that have occurred recently, converting these into a context of viewpoint elements, and retrieving from the knowledge base all traces that match. Each one is then converted back to a sequence of events, the final element of which is interpreted as a prediction of the upcoming event. It is arranged that each viewpoint predicts every possible next event with some non-zero probability. A multiple-viewpoint system derives its power by having many viewpoints working simultaneously on the same prediction problem.

There are two distinct factors that contribute to the prediction of the next event in a piece of music. One is long term: knowledge about the style and the general constraints that it implies. The other is short term: transitory constraints introduced within a particular piece. Ignoring short-term effects would yield poor predictions, since the system would be oblivious to repetition

and patterns within the present piece. To model this problem of the "particular and the general" [1] we introduce, for every viewpoint, a short-term knowledge base that is discarded after each piece is predicted. A viewpoint makes a prediction by combining the independent inferences from its short- and long-term knowledge bases.

For the experiments described below, the machine used a multiple-viewpoint knowledge base learned from a training sample of 95 chorales. A set of 5 chorales was used to measure the entropy of the style. This test set was disjoint from the training set, and the chorales 151 and 61 used in the human experiments were deliberately omitted from the training set. For each of the 220 events in the test chorales, an incomplete piece was shown to the machine, and a probability distribution for all possible continuations was produced. The computational model is general enough to predict all basic types in an event—start-time, duration, pitch, and fermatas. In conformance with the human experiments, the system was instructed to predict only pitch.

4   Experimental Results and Their Musical Interpretation

Chorales 151 and 61 were analyzed in terms of both human and computational models. Average entropies are given, along with the melodies, in Figures 1 and 2. Both "instantaneous" and "windowed-average" entropies are shown. The former (Figures 1a and 2a) is the note-by-note entropy discussed above, averaged over several subjects in the case of the human model. The latter (Figures 1b and 2b) has been smoothed by averaging *over time* as well, and is useful for discerning medium-term trends in the information content of a piece of music. The profiles shown were derived by averaging over seven events of the instantaneous entropy data using a sliding, triangular-shaped window.

For Chorale 151, the entropy of pitch is estimated to be 1.97 bits/event using the human model, while the computational one averages 2.09 bits/event. For Chorale 61, the pitch entropy is 1.53 bits/event for the human model and 1.77 bits/event for the computational one. These results indicate (reassuringly!) that people perform better than the machine at predicting pitch. However, the machine could undoubtedly be tuned to improve its performance, and indeed the instantaneous entropy plots help to indicate just where and how such improvements could be made.

4.1 Chorale 151

Human Model. Instantaneous entropy profiles are most useful for a detailed note-by-note analysis of the music. Consider first the profile in Figure 1a produced by the human model. The chorale exhibits a wave-like entropy, with valleys more or less corresponding to cadence points, and peaks occurring in the middle of phrases. The troughs occur exactly at the cadence points for the first, second, and fourth phrases (notes 7, 15, and 29). The only exception is note 22, where the trough falls at the penultimate note. This perhaps reflects the lack of closure associated with the chorale's only half cadence.

The peak of the wave for phrase 1 occurs at note 4. This is due to the leap of a third which breaks the repetition of the chorale's initial note. Another peak occurs at note 10. Here we expect duplication of the chorale's initial phrase, but the third repetition of the B does not occur and is replaced by a stepwise downward motion. The peak at notes 18 and 19 is also due to a variation on the first phrase. At note 18 there is some uncertainty as to whether the pattern of the first or second phrase will be repeated. Although the second phrase seems to be confirmed, our expectations are thwarted when a new pattern is created at note 19 with the repetition of the A, causing a large increase in entropy.

The biggest peak of the chorale occurs at note 23. An upward leap of a minor seventh is extremely unusual in this style of music, and is doubly unexpected at this point since the leading-tone at note 22 would normally resolve to a G. This leap does make musical sense, though, since the descending stepwise motion to the final tonic is merely displaced by an octave to keep the melody within the range of the soprano. However, when considered on a note-to-note basis, the leap is unexpected and results in the large increase in entropy.

Computational Model. The instantaneous entropy produced by the computational model for Chorale 151 is also shown in Figure 1a. The two profiles often track each other closely, but at several points they diverge substantially. At note 6, for example, the computational model does poorly since it predicts a B with probability 97% and the actual note C with only 2%. Obviously it favors a stepwise resolution of the 7th of the implied dominant harmony to the mediant of the tonic chord (C to B). Because of the model's construction, it cannot look ahead to the upcoming cadence at note 7, where the likely note is B, the third of the tonic triad. At note 7, however, the computational

model does better than the human model, since it predicts the B with 98.6% probability. The descending half-step resolution of the implied 7th is highly favored by the long-term knowledge base at this point, and the choice is reinforced by the short-term knowledge base which has been conditioned to expect the mediant. This illustrates the fact that the computational model is often a very aggressive predictor. People also do well at this point, but they are more cautious and temper their predictions somewhat.

At note 8 the computational model predicts A with 35% probability, C with 29%, and B (the actual note) with 19%. At note 9, A is predicted with 59%, and the actual note B with 15%. Again, the model favors stepwise motion and has not recognized that this phrase begins in the same manner as the initial one. At note 12, B is predicted with 49% and C (the actual note) with 25%. At this point the model seems finally to have realized that repetition of the note B is a statistically significant element of the chorale; however, the note that actually occurs is the model's second choice.

At notes 13 and 14 the computational model outperforms the human one. It predicts the resolution of the 7th of the implied dominant harmony with 79% certainty, and the stepwise motion from the B to the A with 91%. The computational model tends to be an enthusiastic predictor of stepwise motion. At note 15 it does much poorer than humans. It predicts a B with 53% certainty, and G (the actual note) with only 20%. Obviously people are not surprised to see a cadence on the tonic in the middle of the chorale. However, the computational model does not favor this straightforward choice but prefers an imperfect cadence. This is because the model has been conditioned by the short-term knowledge base to favor the note B after so many repetitions.

At notes 16 and 17 the computational model again fails to recognize the tendency of phrases to begin with repeated B's. Instead it favors stepwise motion, predicting at note 16 an A with 33%, G with 20%, and B (the actual note) with 16%; and at note 17 a C with 43% and B (the actual note) with 31% probability. At note 21, this tendency to predict stepwise motion manifests itself as another peak in the entropy. Pitch A is predicted with 37%, B with 21%, and the actual note G with 20%. The computational model cannot look ahead, so it does not know about the high probability of a half cadence at note 22, and thus cannot discern the likelihood of a repeated G to prepare for this event.

Although humans did poorly at predicting note 23, the computational model did somewhat

worse. Its top three predictions were G at 40%, E (one octave below the actual note) at 33%, and F♯ at 17%. All are reasonable choices, although it is apparent that the model cannot explicitly recognize the need for octave displacement of the descending melodic line to avoid exceeding the soprano's range. At note 26 the computational model predicted the B with 96% certainty. It seemed very certain that the 7th of the implied dominant harmony would resolve to the mediant of the tonic, while people were less sure. The peak in entropy at note 28 again points out a major weakness of the computational model—it does not adjust its prediction by looking ahead to the upcoming cadence. At this note the model predicted a B with 35% probability, a G with 32%, and A (the actual note) with 27%. This failure to recognize temporal context also manifests itself in the model's prediction of the final note. Here, the G is predicted at only 69%, a level of confidence that is significantly lower than people's. It is clear that, unlike most musicians, the model does not realize that virtually all melodies end on the tonic.

Windowed-average entropy profile. We turn now to the windowed-average entropy profiles shown in Figure 1b, which can be used to discern medium-term trends in the chorale's information content. The profile for the human subjects shows a simple wavelike structure whose troughs correspond roughly to the cadence points at notes 7, 15, and 29. The cadence at note 22, which occurs just after the high point of the second wave, is an exception. It is interesting to note that the largest excursions in the entropy profile occur in the second half of the piece.

The windowed-average entropy profile produced by the computational model is quite similar in shape to that of the human model, particularly from note 10 onwards. Up to this point the computational model does rather poorly compared to the human one—it is less able to provide good predictions with the limited context that naturally occurs at the beginning of any piece. Note also that even though the shape of the two profiles is similar, the computational model does roughly half a bit per event poorer from note 10 to the end.

4.2 Chorale 61

Human Model. Chorale 61 can be analyzed in a manner similar to Chorale 151. Its instantaneous entropy profiles are shown in Figure 2a. Examining the human-model profile, wave-like motion is again evident, with valleys at each cadence. This is particularly clear at notes 7, 14, 21, 27, 42, 49,

and 57. The cadence at note 34 is the only exception, probably because the expected resolution is Eb whereas a G is heard.

This chorale differs from the other in that we see not one but two peaks in the middle of each phrase, of unequal sizes. The larger peak comes first in phrases 2, 3, 5, 6, 7, and 8, on either the first or the second note following the cadence. This reflects the fact that it is difficult to predict which direction a new phrase will take. In phrase 7 this uncertainty is heightened by the unexpected downward leap of a third at note 44. In phrase 4 the smaller peak comes first, indicating that the cadence's resolution is quite clear, being leading-tone to tonic. The larger peak comes at note 25 with the repetition of the G. Up to this point there has only been one repeated note, so there is a high expectation that all movement will be by step. The minor peak in phrase 6 (note 40) is also due to an unexpected repeated note, and the small peaks at notes 31 and 53 (phrases 5 and 8) can be attributed to the fact that rising motion by step is discontinued in favor of descending stepwise motion.

Computational Model. Figure 2a also shows the instantaneous entropy profile as produced by the computational model. As with Chorale 151, the profiles track each other fairly closely, although they diverge at several points. The computational model consistently beats the human one when it can successfully predict downward stepwise motion. This can be seen at notes 1–3, 8–12, 18–19, 30–33, 48–49, and 52–55. The computational model's advantage at these events is small—usually less than half a bit.

This aggressive tendency of the computational model to predict downward stepwise motion works against it at several other points. For example, at notes 13, 28, 36, 40, and 56 it consistently predicts such a motion but it is repeated notes that actually occur. Since the predictions are wrong, the computational model shows entropy peaks at these points. The result is disastrous at note 28, since the model predicts an F with 99.7% certainty, leaving very little probability space for the G that actually occurs.

The model's propensity to predict downward stepwise motion also creates entropy peaks at notes 15–16 and notes 44–45. Here, intervals larger than a second occur. One should observe that people also had difficulty predicting these notes, but still performed about 2 or 3 bits better than the computational model. At notes 6 and 26 we see motion from G to Ab, which to the machine is

11

relatively surprising, again partly because of the model's propensity to predict downward stepwise motion.

At note 38 the chorale modulates to the dominant with the introduction of the A♮. To people this is not at all surprising, as indicated by the slight decrease in entropy at this point. To the computational model, however, a peak of entropy occurs since an F is predicted with 44% certainty, and the A♮ with only 20%. At note 47 the chorale modulates back to the tonic with the re-introduction of the A♭. This is somewhat surprising to the human subjects, as shown by the small peak in entropy, but the computational model has even more difficulty since it predicts A♮ with 57% certainty and the A♭ with only 26%. The humans had no difficulty in recognizing that the modulation back to the tonic was permanent, whereas the computational model predicted another A♮ at note 51 with a confidence of 31%, F with 25%, and A♭ (the note that actually occurs) with only 23%.

The inability of the computational model to look ahead for cadence points accounts for the sharp peak in entropy at notes 20–22. Here, the half cadence is not at all surprising to humans, but presents difficulty for the model. The lack of goal orientation and the inability to recognize temporal position works against it at this point, and also at the penultimate note of the chorale, where people had no difficulty recognizing the need for a repeated note to prepare for the perfect authentic cadence.

Windowed-average entropy profile. Figure 2b shows the windowed-average entropy profile for Chorale 61 as generated by the human and computational models. Recall that the profile produced by the human model shows medium-term information trends. Like Chorale 151, there is a tendency for the troughs of the waves to fall more or less at cadence points. This is true for the cadences at notes 7, 14, 21, 34, 42, 49, and 57, but not at note 27. It is interesting that—as in Chorale 151—the largest excursions in entropy occur in the second half of the piece. One can perhaps speculate that to avoid monotony there is a need to inject some uncertainty toward the end, just before the final resolution.

The profile produced by the computational model tracks the human one quite closely, and their shapes are remarkably similar, although the former model does somewhat poorer at most events. This is particularly striking at notes 16, 27, and 44, and is explained by the effect of the model's

poor predictions at these points on the windowed average. However, what is encouraging is that the computational model can, with some tuning, closely approximate the human model's predictive ability, and thus produce accurate windowed-average entropy profiles. This is important because these profiles can be used to characterize medium-term information trends in the music.

5 Improving the Computational Model

It seems possible that a machine will eventually outperform people at the chorale guessing game. This section outlines some potential improvements to the multiple-viewpoint model that may achieve this goal. The improvements are divided into two classes: slight parameter adjustments to the existing model, and structural changes to the modeling and representation framework.

One very simple adjustment that could be made is to increase the corpus of example chorales used to create the model. Experiments have indicated that performance improves as more training chorales are seen. Only 95 of the many existing Bach chorales were used. A second area for improvement is to reconsider the methods used for selecting and combining viewpoints. As we have seen, the system is at times an overly aggressive predictor. Its predictions could be tempered by using a different technique for combining the predictions from individual viewpoints. Furthermore, it should be possible to develop a more principled way of selecting individual viewpoints from the huge number of possibilities available. Finally, it would be profitable to explore different ways of combining predictions from the short- and long-term viewpoint models. For example, the relative weighting of the short-term model could be increased as a chorale progresses. The wide discrepancy between human and machine performance in the opening notes of Chorale 151 (Figure 1b) highlights the weakness of having a fixed weighting.

Several structural changes to the framework should be investigated. Our computational model has some significant shortcomings. It does not incorporate *a priori* musical knowledge (except that implicit in the selection of viewpoints). It does not plan, or learn from failure. It uses no form of hierarchical representation or control over its predictions. Here are some ways that such features might be introduced.

First, one could incorporate some kind of constraint mechanism that codifies inviolable rules of chorale construction. The current prediction system was designed to be *nonexclusive*, predicting

13

all possible events with some non-zero probability. Some form of "hard" constraint could be used to reduce viewpoint prediction sets, opening up probability space for unconstrained events. This would extend the system from a pure learning system to a hybrid constraint-based learning system.

A second area for improvement is in rhythm and lookahead. The system's current structure prohibits it from looking ahead to inspect the rhythmic skeleton: it only takes into account the past context, not the future. People are free to scan the complete rhythmic surface, gaining useful *a priori* constraints on pitches. This is quite striking at cadences, and in particular at the final cadence.

Third, it is likely that different viewpoints change their relative importance as a piece progresses. To accommodate this, a meta-model could predict which viewpoints should be invoked at various stages. Such changes in perspective could perhaps be learned by inspecting and reasoning about the entropy profile. This kind of introspection about the performance of the model could be called "meta-reasoning." As an example, the modulation in Chorale 61 indicates that people come to expect excitement or variety to be encountered as a piece progresses.

Learning systems often avoid overgeneralization by using negative examples. The utility of negative examples here is not clear because our learning algorithm is probabilistic, and implicitly searches a *specialization* hierarchy rather than the more usual *generalization hierarchy*. However, negative examples could be used to adjust the statistics given to various traces. Alternatively, a "negative" viewpoint could be used to bias predictions.

6 Conclusions

This is the first detailed, quantitative comparison of the performance of a computational model of music with the human listener. The basis for comparison is predictive ability, measured in terms of entropy. To investigate and compare both computational and human models required the use of novel techniques. For the former, the development of the multiple-viewpoint scheme and the use of machine learning are both significant innovations that account, in large part, for the success achieved. In order to elicit human models, the methodology of the guessing game was instrumental in gaining quantitative access to people's predictions.

Probably the most striking feature of the experiments was the similarity in performance of

the two systems. While we have naturally tended to focus on and discuss the differences, even cursory examination of Figures 1 and 2 indicates remarkable similarities, in both instantaneous entropy profile and windowed-average profile, between the computational and human models. By and large, both encountered difficulties at precisely the same points in the chorales: the differences are in the amount of difficulty experienced rather than where it occurs. This indicates that it is fundamental musical characteristics that are being measured, rather than some artifact of the measuring instruments.

Finally, people still out-perform the computational model in predicting chorales. However, there is a great deal of room for improvement in the latter, and these experiments indicate the areas in which further research is needed. It does not seem unlikely that machines will eventually outclass people in predicting a restricted and well-defined genre of music.
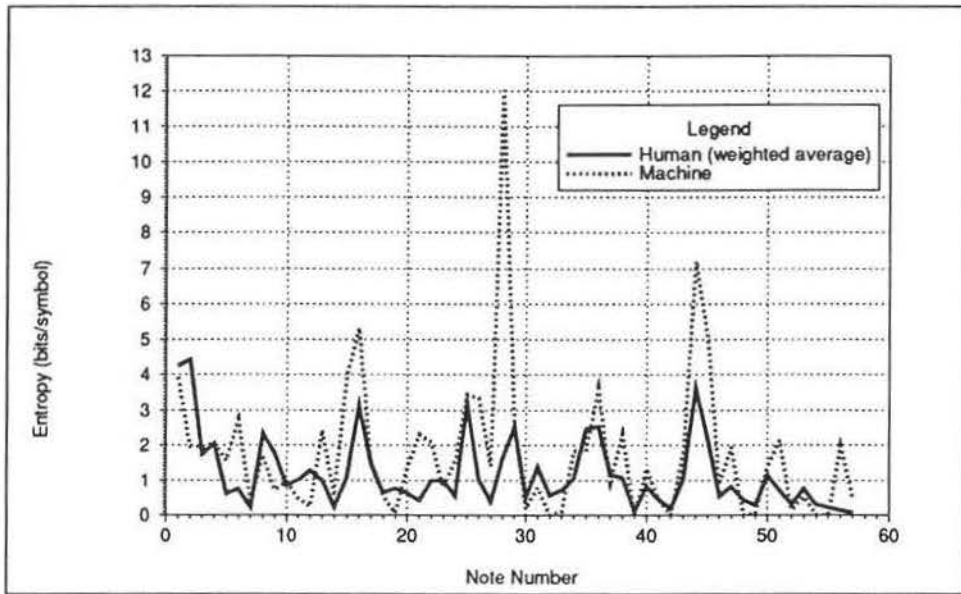
Acknowledgements

References

[1] Brown, M. and Dempster, D. J., "The scientific image of music theory," *Journal of Music Theory* **33**, No. 1 (Spring, 1989).

[2] Conklin, D. and Witten, I.H. *Prediction and Entropy of Music* (Technical Report 91/457/41, Department of Computer Science, University of Calgary, 1991).

[3] Cover, T.M. and King, R.C., "A Convergent Gambling Estimate of the Entropy of English," *IEEE Transactions on Information Theory* **24**, No. 4, 413–421 (July, 1978).

[4] Manzara, L.C., Witten, I.H., and James, M., "On the Entropy of Music: An Experiment with Bach Chorale Melodies," *Leonardo*, in press. (Also available as Technical Report 91/456/40, Department of Computer Science, University of Calgary, 1991).

[5] Shannon, C.E., "Prediction and Entropy of Printed English," *Bell System Technical Journal* **30**, 50–64 (January, 1951).
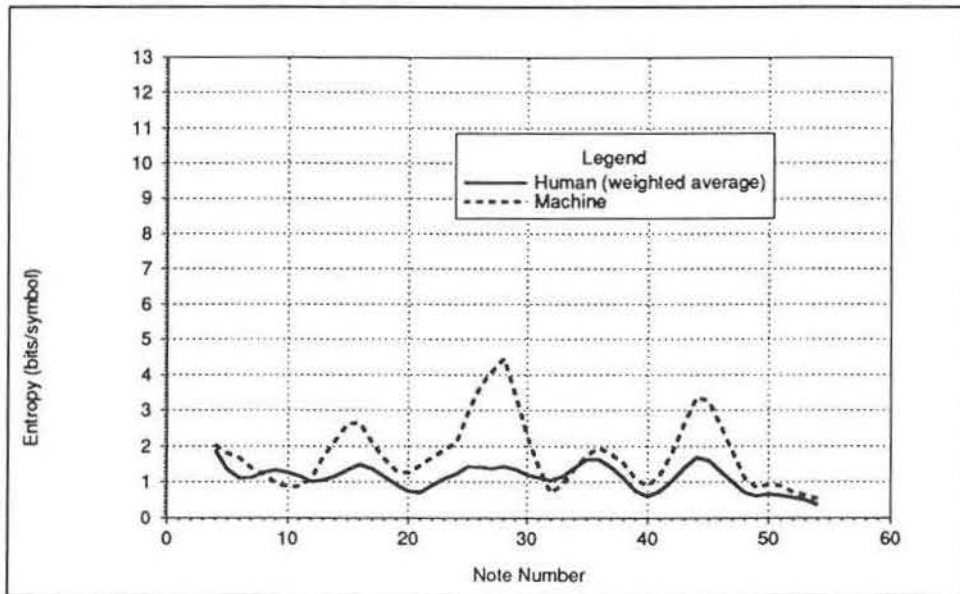
List of Figures

(a)

(b)

(c)

Figure 2