

1 Introduction

Entropy is a measure of disorder, or unpredictability. In a communication system, entropy measures the amount of information that is contained in a message. As messages become more disordered, that is, less predictable, the more difficult it is to transmit them efficiently—in other words, more symbols are needed to represent each one. Conversely, messages that are highly ordered, or predictable, can be coded into fewer symbols.

Many systems can be viewed in terms of information theory: text messages, communication lines, and spoken languages, to name a few. Music, too, can be productively analyzed from this vantage point. Most musicians have an instinctive feel for the amount of information communicated by the music they play and hear. For example, few perceptive listeners would disagree that the compositions of Berg or Xenakis are much more information-laden than those of Mozart. If perceptual complexity were measured objectively in terms of entropy, we could make meaningful stylistic comparisons between individual pieces of music, and, perhaps, between composers and across genres.

Furthermore, the amount of information conveyed by a particular piece of music varies as the piece progresses. We feel this instinctively, for example, in the stretto section of a fugue or in the development section of a sonata. The entropy profile of a piece of music could be viewed on a note-to-note basis, or from phrase to phrase, or from section to section, giving the music theorist another tool to characterize the formal construction of particular compositions.

The entropy of a sequence of events is measured with respect to a “model” of possible continuations. The model predicts upcoming events by assigning probabilities to the set of all possible next events. If the actual next event is given a high probability by the model, it contributes only a little to the entropy of the sequence. If it receives a very low probability, it contributes a great deal. Thus, given a model for a certain genre of music, the entropy profile of a particular piece within it can be calculated from the probabilities that the model assigns to the events that occur in the piece.

This paper describes an experiment that was designed to investigate human listeners’ models of music. We sought an accurate estimate of the entropy of individual pieces within a particular genre of music. We were interested in generalizing the estimate to the genre as a whole, and specializing it to provide detailed profiles of individual pieces. Such an estimate could find use in comparative stylistic analyses of music, both within the genre and, ultimately, across stylistic boundaries.

Apart from the intrinsic interest of this enterprise, we hoped to establish a systematic method for finding objective correlates to elements of musical structure and style that might lay the groundwork for a method of musical analysis based upon objective, repeatable, experiments using actual musicians. We also wanted to provide a benchmark against which the performance of other predictive models of music could be assessed. For example, Conklin [1] describes an adaptive scheme for predicting music based on techniques of machine learning, and applies it to certain compositions. Unfortunately, it is hard to judge the success of this in human terms since it is not known how well people can predict the same compositions.

The music used in the experiments comprised selected melodies from the Chorale harmonizations of J.S. Bach. This music constitutes a well-recognized genre and forms a standard subject of study in most university-level music theory classes. Trained musicians are generally intimately familiar with the chorales. Moreover, they are readily available in machine-readable form, so a large database could easily be assembled for the experiments. Note, incidentally, that these melodies were not composed by Bach, but have been part of the Lutheran Church music tradition since the 16th century.

The methodology we use is suitable for investigating a wide variety of musical phenomena. The experiments we have performed so far, being the first in a planned series, focus on the pitch of the melodies. We thought it best to concentrate on a single parameter of music, and to begin by disregarding the complex interdependencies between harmony, pitch, and rhythm. Subsequent experiments will explore other areas—in particular, the effect of harmonic context upon the entropy profile for pitch.

This paper is organized as follows. The first section sketches the historical and theoretical background for the experiments, which derive from the pioneering work of Shannon [2, 3] and Cover & King [4] on estimating the entropy of printed English. Next we describe the experiment itself. It is performed using a program for the Apple Macintosh computer called CHORALE CASINO, which is freely available from the authors. We outline how the program works and how it is used in our experiments. The third section presents experimental results, and these are evaluated and interpreted in the fourth. Finally we draw some general conclusions and summarize directions for further work. An Appendix sets out in more detail the development of information theory and entropy estimation techniques, and includes a mathematical account of relevant aspects of the

theory.

2 Background

Shannon [5] was the first to investigate the entropy of messages in communication systems. He sought to quantify the amount of redundancy in the English language—other researchers have since applied his methods to a great variety of other natural languages [6]. His early studies were strictly statistical, using letter and word frequency tables derived from samples of text. However, due to the theoretical limits of the technique, these so-called “N-gram analyses” yielded estimates that were deemed to be poor indicators of the true entropy of the source text (see the Appendix for details).

2.1 The guessing game

Realizing the limitations of N-gram analysis, Shannon took a new tack altogether. He recognized that every person has a sophisticated model of language in their head. This model encompasses word structure, syntax, semantics, idioms and style, not to mention the subject matter of the text itself. It allows people to predict effortlessly the letter or word that follows a given context, usually with high accuracy. It also enables readers to spot subtle errors of grammar and spelling, and to make the proper correction, sometimes unconsciously, and usually with great speed.

Shannon used people’s predictive ability to derive an entropy estimate of English. The procedure he followed 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 character;
3. If the guess is wrong, have the subject guess again;
4. Repeat until the guess is correct;
5. Have the subject guess the next character in the same manner, continuing until the entire passage has been completed.

The number of guesses taken by the subject for each character of the text is recorded, and is used to derive an entropy estimate.

Shannon used this technique with 100 short samples of text from the book *Jefferson the Virginian* by Dumas Malone. He concluded that the entropy for this particular source of printed English lies between 0.6 and 1.3 bits per character [7]. The problem with this result is the wide gap between the lower and upper bounds, which is due to the limitations of the guessing technique. For example, our expectation that the letter “u” will follow “q” is much stronger than our expectation that “a” will follow “r”; yet the two events appear to convey the same information if they are guessed in the same number of attempts. Obviously, some valuable information about the text is being lost.

2.2 The gambling game

Cover & King [8] improved upon Shannon’s technique by incorporating gambling into the guessing procedure. Each subject is given an initial capital amount of $S_0 = 1$. At each stage n , if the player guesses correctly his capital S_n is set to $27pS_{n-1}$, where p is the proportion of capital the subject chose to wager and S_{n-1} is the capital accrued from the previous stage. If the player guesses incorrectly, the capital declines by the proportion bet. The number 27 represents fair odds because there are 27 symbols to choose from (26 letters plus a space).

It can be shown that the entropy of text can be estimated reliably from the capital accumulated by a subject [9]. Mathematically, it is expressed as:

$$H = \lim_{n \rightarrow \infty} \left[\log_2 27 - \frac{1}{n} \log_2 S_n \right], \quad (1)$$

where H is the entropy of the text in bits per symbol, n is the current stage (i.e. the number of the letter being guessed), and S_n is the capital accumulated by the subject after n stages. Any practical estimate must, of course, be based on a finite number of guesses, and its accuracy will increase as the number of letters grows until, in the limit, the “true” information content of the text is obtained.

The gambling procedure elicits from the subject an intuitive probability estimate for the next symbol to be guessed. If a particular symbol is considered very likely to occur, good subjects will wager a large proportion of their capital on that choice. Conversely, in situations where subjects are less sure of the symbol they will bet a smaller amount. This scheme not only rewards the correct guess, but rewards accurate estimates of the symbol’s probability—gamblers will recognize it as proportional betting. To obtain the best possible results, subjects are pitted against each other in a tournament competition, with a reward to the overall winner.

An important practical advantage of the scheme is that subjects are required to supply a probability estimate only for the symbol they feel is most likely. If that turns out to be incorrect, they must supply a probability for the next most likely symbol, and so on until the correct one is chosen. This sequential betting procedure reduces the time required for the subject to proceed through the text because probability values need not be supplied for every possible symbol at every stage of the experiment.

2.3 Results from the gambling game

One way of combining the results from several subjects in this experiment is simply to average the capital that accrues to each one and calculate the entropy from it using equation (1). This is called the “average capital estimate.” However, the best subject’s entropy value is more telling because it reflects the most sophisticated model of the language—we are not particularly concerned with players who have poor models. While the “best subject estimate” is useful, it is unreliable. It may be optimistic because with a large number of players there is a chance that the best one does well simply by luck. It may be pessimistic because the best overall player may suffer occasional lapses of judgement—a joint decision by all players could do better than any one individual. Both of these problems can be solved by using a “committee gambling estimate,” which simply uses a weighted average of the individual accrued capitals, with weights chosen in proportion to the subject’s relative success in the tournament. This can sometimes give an even better entropy figure than the best subject [10].

Using the same text as in Shannon’s experiment, Cover & King [11] arrived at an average capital estimate of 1.34 bits/symbol, a best subject estimate of 1.29 bits/symbol, and a committee gambling estimate of 1.25 bits/symbol. They were careful to note that each source of printed English will have its own characteristic entropy, and these estimates are valid for this particular text only.

2.4 Entropy of Music

The pioneering work of Shannon soon led to the application of information theory to music. Some researchers, such as Meyer [12], were interested in the aesthetic and philosophic implications of the theory when applied to music. Others studied specific musical examples. For example, Hiller & Fuller [13] analyzed four sonata expositions statistically, and derived a variety of “contours

of information fluctuation.” From these they drew a number of conclusions about the style of the composers, and made meaningful comparisons between the works. Other similar studies soon followed [14, 15]. However, all these analyses used N-gram techniques, with the attendant statistical uncertainty, and say little about people’s models of music.

3 Description of the experiment

3.1 The CHORALE CASINO program

The CHORALE CASINO program was written to facilitate the investigation of human performance in music prediction, using the gambling methodology. The program has a built-in database of one hundred melodies, derived from the chorale harmonizations of J.S. Bach. The experimenter can choose any of these melodies simply by entering a security password and the desired chorale number.

The program is structured as a game, and records each player’s score and rank within a tournament-style competition. The result of every guess, and the amount bet, are recorded in the subject’s log file, as is the time taken to make the bet. The program also calculates and stores the entropy at each note of the chorale.

When CHORALE CASINO begins, players are presented with a registration panel which asks them to enter their name and years of experience in music—both practice and theory. The program then presents a series of tutorial panels explaining how the subject is to play the game. Next, the current rankings in the tournament are displayed, giving the names and scores of the top five players so far.

The game proper now begins, and the main display is shown on the screen. Figure 1 shows the situation partway through a chorale. It depicts the rhythmic skeleton of the music, with all notes still to be chosen marked in gray in a neutral position—the middle line of the staff. Bar lines are shown, as is the key signature, time signature, and fermata signs; however, repeat signs are omitted. If the entire chorale is too big to fit on the monitor, the player can move to any part of it by pointing (with a mouse) to the scroll bar and clicking on it.

The player chooses a pitch to bet on by manipulating what we call the “note slider.” This consists of an up-arrow button above the current note and a down-arrow button below it. When the former is clicked using the mouse, the note displayed on the staff rises chromatically. When the

latter is selected the note falls chromatically. In this fashion the player can choose any pitch from the chromatic scale between middle C and G at the top of the staff. However, if a note has already been unsuccessfully bet upon, the slider will skip over that particular note, preventing the player from the pointless exercise of betting on it again.

After the note has been identified correctly, the slider advances to the next one. The correctly guessed melody remains on the screen, while notes still to be chosen are displayed as grayed notes in neutral position. The player can hear the melody up to and including the current point by selecting the “hear it” item in the menu bar at the top of the screen, which causes the notes to be synthesized by built-in hardware and played on the computer’s speaker. When the game begins the note slider is placed over the chorale’s first note, and the music is guessed in order from left to right.

Players must bet on each pitch they choose. When they have manipulated the note slider until they are satisfied with the note selected, they push the “Note OK” button attached to it. This immediately brings up the bet panel illustrated in Figure 2. It consists of the “bet slider” to the left, and four numerical displays to the right. The player chooses the proportion of capital to bet by moving the slider up and down with the mouse. As it moves, the numerical displays are automatically updated. Two of these show the amount bet, one as a proportion of capital and the other as an absolute dollar figure. The remaining two show the potential total capital, one for a losing bet and the other for a winner. When the player is satisfied with the amount to be wagered, the “Bet amount shown” button must be pressed. If the pitch selected is correct, a panel is displayed telling the player that they have won, along with the updated capital amount; and the note slider is automatically advanced to the next note of the chorale. If the pitch is incorrect, the player is informed of the capital remaining after the loss. The note slider will not advance in this case, forcing another choice of pitch at the same point in the chorale.

The player continues to choose, and bet on, each note until the whole piece has been completed. At the end the final capital amount is displayed, along with the player’s rank in the tournament.

3.2 Experimental trials

A tournament was organized to collect experimental data. We wanted an accurate estimate of the entropy of particular pieces of music, and this is best achieved when a strong sense of competition is fostered among players. Contestants were grouped into three categories: expert, intermediate,

and novice. A novice was defined as any person with the basic ability to read music, while an intermediate player must have had at least two years of formal music theory classes. An expert was defined as any person with advanced music degrees and professional experience.

Five contestants in each category competed against each other in the first round of the tournament. The top two players from each category advanced to the playoff round. As an incentive to perform well, a prize of \$100 was awarded to each winner of the playoff round in the novice and intermediate categories. The melody employed for the first round of the tournament was taken from Chorale No. 151 of J.S. Bach's *371 Four-Part Chorales* (Edition Breitkopf). For the second round, Chorale No. 61 was used. Care was taken that no contestant knew in advance which chorale melody would be chosen.

The experiments were performed using the CHORALE CASINO program. After a short briefing, the contestants were left alone at the computer to complete the game without any interruptions. No time limit was imposed for completion, though some players took as long as 64 minutes while others finished Chorale 151 in only 17 minutes.

4 Experimental results

The cumulative entropy profile for the melody of Chorale 151 is shown in Figure 3a, which plots the cumulative average entropy in bits per symbol for each note. This was calculated by:

$$H = \lim_{n \rightarrow \infty} \left[\log_2 20 - \frac{1}{n} \log_2 S_n \right], \quad (2)$$

where n is the current stage and S_n the capital accumulated so far by the subject. The number 20 represents fair odds because this is the number of chromatic pitches to choose from (middle C to G above the staff). Three profiles are given: the average capital estimate, the weighted average estimate, and the best subject estimate. The first is calculated at each note by averaging the accumulated capital of all 15 subjects and then applying the above formula. The second is obtained similarly, except that each subject's capital is weighted by the factor $capital/total$, where $total$ is the sum of all subjects' accumulated capital at that particular stage. The third is formed by taking the highest accumulated capital of any subject at each note. Note that the best subject and weighted average estimates converge fairly quickly to the same value.

We conclude that the average capital estimate for the entropy of this chorale is 2.086 bits per symbol. The weighted average estimate is 1.982 bits/symbol, and the best subject estimate is 1.974 bits/symbol.

Since the cumulative entropy estimates are running averages, the profiles tends to settle down as the number of notes increases, damping any short-term variations in entropy. In order to show these variations, instantaneous entropy profiles were also obtained, and are shown in Figure 3b. Instantaneous entropy at each note is calculated by:

$$H = \log_2 20 - \log_2 S , \quad (3)$$

where H is the entropy of the current pitch in bits per symbol and S the capital won by the subject on this note. This formula is derived from equation (2) by setting n to 1, which is tantamount to assuming that subjects start out with a capital of 1 at each note. The average capital, weighted average, and best subject estimates are calculated in the same manner as for cumulative entropy, except, of course, that the capital amounts are not cumulative.

Of the three estimates of instantaneous entropy, the weighted average is probably the most useful. The best subject estimate really just shows that at least one of the 15 subjects managed to score much better than all the others, with the result that several flat spots appear in the graph. On the other hand, the average capital estimate can be unduly influenced by a single poor bettor on a particular note, with the result that the profile may show a peak in entropy where one should not really exist. The weighted average is the best compromise because it minimizes the effect of poor betting but does not unduly reward the occasional lucky bet.

The cumulative entropy profiles for the melody of Chorale 61 are shown in Figure 4a. These are calculated in exactly the same way as for Chorale 151, except that only 6 subjects were used in the playoff round. The average capital estimate for the entropy of this chorale is 1.575 bits per symbol, while the best subject and weighted average estimates are both 1.529 bits/symbol. The instantaneous entropy profiles for the chorale are given in Figure 4b. They, too, are calculated in the same way as for Chorale 151, and again only 6 subjects were used.

Surprisingly, there was little difference in the results between the three categories of subjects. Once extreme results were discarded, we found that each group averaged about the same. In fact, the best performer was one whom we had classified as a “novice”! This may speak more about

shortcomings in our categorization and the positive effects of careful betting practices than it does about musical models.

5 Evaluation and interpretation

5.1 Lower bounds estimate of the entropy

According to the best subject estimates, the entropy of pitch in Chorales 151 and 61 is 1.974 and 1.529 bits per symbol respectively. Let us consider how closely these figures might approximate the “true” entropy of the chorales. The experiment is designed to evoke each subject’s inherent statistical model of music to derive the entropy profiles. It is assumed that subjects have a good understanding of proportional betting—for the experiment rewards those who bet according to the probabilities they attribute to the pitches. Thus, over and above their musical understanding and experience, subjects’ gambling skills are involved. In some cases we found that these skills were quite poor and acted against an obviously solid grasp of the music. On these grounds alone, we expect that the entropy estimates would be lowered if the subjects were shown how to wager effectively and allowed to practice their betting.

As with any game, participants can be expected to improve their performance with practice. With each new round of the tournament, subjects became more comfortable with the task, not only in their betting technique, but also in terms of thinking about musical processes in a completely new way. As subjects become more proficient at turning their musical understanding towards playing the game, we expect a further drop in the entropy estimate.

Another factor that may have artificially increased the estimate is the use of forced sequential guessing. In these experiments, subjects were constrained to guess the pitch of each note in order, from left to right. Although we often think of music as a sequential entity progressing through time, at a deeper structural level its organization is most definitely non-sequential. Future experiments will explore this issue by allowing subjects to choose notes to work on in any order. Preliminary results indicate that the corresponding entropy estimate will drop noticeably.

In principle, the estimates we derive converge to the “true” entropy as the context increases to infinity (see equation (1)). The chorales are limited in size, so the asymptote will never be reached. It may be that entropy estimates for larger chorales tend to be lower simply because of this fact. The

results above seem to support this, but many more trials are needed to yield conclusive evidence. Alternatively, it may be that the estimate converges quickly and the entropy profiles accurately reflect the changing information content of the music beyond the context of a few notes.

Notwithstanding these factors, it seems fair to say that 1.974 and 1.529 bits per note are reasonably accurate entropy estimates for pitch in the two chorales, although they probably err on the high side. Unfortunately it is not possible to extrapolate these into an overall entropy figure for the chorale genre—just two estimates are quite insufficient to support any degree of confidence in the result. A much larger number would be required to generalize to the class of pieces in a statistically valid manner, and we will need to collect more data before we can produce an average entropy estimate for the chorales. Such a figure would be useful as a basis for comparing the chorale melodies with other genres of music. However, it provides only a gross comparison between styles, and the individual entropy estimates are more interesting because they can be used for detailed analyses of particular pieces of music.

5.2 Correlation with stylistic analysis of the music

The instantaneous entropy profile can be used for a detailed note-by-note analysis of the melody of the chorale. There are four phrases in Chorale 151 (see Figure 3c), and they are marked by fermata signs in Figures 3a and 3b. Using the weighted average estimate, it is apparent that the “valleys” in the instantaneous entropy profile of Figure 3b tend to correspond to the ultimate notes in each cadence. This is particularly clear in the cadences at notes 7, 15, and 29. In the third cadence, the dip in the profile corresponds to the penultimate note, and there is a slight rise in entropy when the cadence finally appears at note 22. The difference may be due to the fact that this cadence is the only one which is not a V–I cadence. It can be considered a half cadence if the harmonization of Bach is ignored (and recall that the subjects are given only the melody), and the slight rise in entropy seems to correspond to the consequent lack of closure.

The chorale exhibits a wave-like entropy profile, with troughs corresponding more or less to cadence points and peaks occurring at the middle of phrases. For example, the peak in the first phrase occurs on note 4. Here a leap of a third breaks the repetition of the chorale’s initial note. At note 10 we see another sharp peak, due to the fact that the expected duplication of the beginning of the first phrase does not occur, but instead is replaced by a stepwise downward motion. The peak

at notes 18 and 19 is again due to the variation on the beginning phrase. At note 18 there is some uncertainty as to whether the pattern of the first or second phrase will be repeated exactly. When the second pattern seems to be confirmed, a new pattern is created at note 19 with the repetition of the A, causing a large increase in entropy.

The largest peak of the chorale occurs at note 23. The leap of a minor seventh is very unusual in this style of music, and is even more unexpected in this case because the leading-tone at note 22 would normally resolve to a G. In retrospect, this leap does make musical sense since the descending stepwise motion to the tonic is displaced by an octave to keep the melody within the soprano's range. However, on a note-to-note basis the leap is very unexpected, and results in the large rise in entropy.

The cumulative entropy profile of Figure 3a is most useful for discerning long-term trends in the chorale's information content. The part of the profile over the first 5 or 6 notes should be ignored since there is insufficient context to establish a long-term tendency. In contrast, the second and third phrases show a definite steady decrease in entropy. Whether this is due to the convergent nature of the estimate or the musical nature of the melody is open to some debate. It is clear, however, that enough context has been built up to allow the subjects' predictive ability to grow progressively stronger through these two phrases. At note 23 the leap of the seventh shows up very clearly even on this graph as a relatively sharp peak, while the rather predictable ending is reflected in a fairly steep decline in entropy. In summary, one can characterize the first phrase as uncertain due to the lack of context, phrases two and three as increasingly predictable, and phrase four as starting with a jolt but ending in a logical manner.

Chorale 61 can be analyzed in a similar manner. As in Chorale 151, we see wave-like motion in the instantaneous entropy (Figure 4b), with troughs at each cadence point. This is particularly clear at the cadences at notes 7, 14, 21, 27, 42, 49, and 57. The one at note 34 is the only exception, presumably because the expected resolution is Eb, whereas a G is heard.

This chorale is interesting because we see not one but two peaks, of unequal size, in the middle of each phrase. In phrases 2, 3, 5, 6, 7, and 8 the larger peak comes first, either on the first or the second note after the cadence point. This reflects that fact that after a cadence, it becomes difficult to predict what direction the new phrase will take. In phrase 7 this is compounded by the unexpected downward leap of a third on note 44. In phrase 4 the smaller peak occurs first, indicating that the resolution of the cadence is quite clear, since it is leading-tone to tonic. The large peak comes on a

repetition of the G at note 25. Up to this point there has been only one repeated note, so subjects have a high expectation that all movement is by step. The minor peak in phrase 6 (note 40) is also due to an unexpected repeated note, while the minor peaks in phrases 5 and 8 (notes 31 and 53) can be attributed to the fact that the rising stepwise motion is discontinued in favor of descending motion by step.

Figure 4a shows the cumulative entropy profile for Chorale 61. As in the other chorale, the profile for the first phrase should be ignored due to lack of context. Phrases 2 through 5 show a more or less steady decline in the entropy. What is interesting are the two slight rises in phrases 6 and 7 (at notes 35–38 and 43–44). These correspond with the modulation to the dominant in the former and the unexpected leap at note 44 in the latter. The final phrase continues the steady decline in entropy started in phrase 7.

There is a remarkable similarity in the long-term trends of these two chorales. Both exhibit uncertain entropy profiles in the first phrase because of a lack of context. The next several phrases show a decreasing entropy, reflecting the fact that the growing context supports more confident prediction. Finally, both chorales have a rise of entropy near the end of the piece, followed by steep declines to the final cadence. Part of what seems to make a satisfactory melodic profile is a pattern of uncertainty, increasing stability, uncertainty, and final resolution.

5.3 Comparison with algorithmic models of prediction

As mentioned in the introduction, we wanted to establish a benchmark against which a predictive model based on machine learning could be judged. A detailed comparison is left to another paper [16]. However, we briefly note that the derived entropy profiles are remarkably similar, though there are important exceptions due to the nature of the algorithmic model. The human subjects slightly outperformed the machine-based system, although we expect that the model can be tuned to more closely approximate, or possibly surpass, these results.

6 Future experiments

The experiments performed above lead in a number of new directions. We first want to investigate the effect of non-sequential guessing in the gambling methodology. The CHORALE CASINO program

has already been modified to allow this, and preliminary results indicate that the entropy estimates will drop significantly. A second planned experiment will provide subjects with the context of the three other voices of the chorale when guessing the pitch of the melody. The data obtained from this study will help us to improve the algorithmic model when it deals with multi-part music. Finally, the gambling methodology will be applied to parameters other than pitch—rhythm and harmony, in particular, should prove to be rewarding areas of investigation.

7 Conclusions

The computer program CHORALE CASINO was used to administer experiments designed to elicit predictive probabilities of music from subjects. From this data we derived entropy estimates for pitch of 1.974 and 1.529 bits/symbol for Chorales 151 and 61 respectively. More importantly, cumulative and instantaneous entropy profiles for each of the chorales were derived. These proved useful in musical analyses of the pieces, and helped to characterize the short and long term structure of the music. If enough data is collected, such studies can be used to characterize a composer's style in general, and to provide an objective, scientifically repeatable, means of comparative analysis. The results of these studies will also be useful when constructing and tuning machine-based predictive models of music.

Acknowledgements

We would like to acknowledge the key role of Darrell Conklin, who helped develop the idea of performing experiments on people's ability to predict music. Many thanks go to all the subjects who participated in the experimental trials. Geoffrey Falk won in the novice category, Windsor Viney in the intermediate category, and Dr. William Jordan in the expert category. Special recognition is due to Geoffrey Falk, who, despite being a "novice," outperformed all other contestants in both rounds of the tournament. This work is supported by a grant from the National Sciences and Engineering Research Council of Canada.

References

- [1] D. Conklin, *Prediction and Entropy of Music* (MSc Thesis, Department of Computer Science, University of Calgary, 1990).
- [2] C.E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (Urbana, Illinois: University of Illinois Press, 1949).
- [3] C.E. Shannon, "Prediction and Entropy of Printed English", *Bell System Technical Journal* **30**, 50–64 (January, 1951).
- [4] T.M. Cover and R.C. King, "A Convergent Gambling Estimate of the Entropy of English", *IEEE Transactions on Information Theory* **24**, No. 4, 413–421 (July, 1978).
- [5] Shannon and Weaver [2], Shannon [3].
- [6] I.H. Witten and T.C. Bell, "Source Models for Natural Language Text", *International Journal of Man-Machine Studies* **32**, No. 5, 545–579 (May, 1990).
- [7] A bit is a binary digit, with the value 0 or 1. It is the smallest unit used for quantifying amounts of information.
- [8] Cover and King [4].
- [9] Cover and King [4], p. 415.
- [10] Cover and King [4], p. 418–19.
- [11] Cover and King [4], p. 419.
- [12] L.B. Meyer, "Meaning in Music and Information Theory", *Journal of Aesthetics and Art Criticism* **15**, No. 4, 412–424 (June, 1957).
- [13] L. Hiller and C. Bean, "Information Theory Analysis of Four Sonata Expositions", *Journal of Music Theory* **10**, No. 1, 96–137 (Spring, 1966).
- [14] L. Hiller and R. Fuller, "Structure and Information in Webern's *Symphonie, Op. 21*", *Journal of Music Theory* **11**, No. 1, 60–115 (Spring, 1967).

- [15] L. Knopoff and W. Hutchinson, “Entropy as a Measure of Style: The Influence of Sample Length”, *Journal of Music Theory* 27, No. 1, 75–97 (Spring, 1970).
- [16] L.C. Manzara, D. Conklin, and I.H. Witten, “A Comparison Between Human and Machine-Based Models of Music Prediction”. In preparation.
- [17] R.V.L. Hartley, “Transmission of Information”, *Bell System Technical Journal* 7, 535–563 (July, 1928).
- [18] For this reason, some people prefer the term “communication theory” to “information theory”.
- [19] Shannon and Weaver [2].
- [20] Shannon [3].
- [21] Shannon [3], p. 50.
- [22] Shannon [3], p. 50.
- [23] Shannon [3], p. 54.
- [24] H. Kucera and W.N. Francis, *Computational Analysis of Present-Day American English* (Providence, RI: Brown University Press, 1967).

Appendix

Information theory

The foundation of modern information theory was established by Hartley in a paper published in 1928 [17]. He stated that any communication system could be examined independently of the people who send or receive the messages, and that it was amenable to objective measurement. Messages sent through a communication system are constructed as a sequence of physical signals. Each sequence can be described as a particular arrangement of arbitrary, discrete symbols. When characterizing a communication system in objective terms, it is not necessary to consider what these symbols mean, but only how many different ones there are and how quickly they can be transmitted through the system. This knowledge can help to estimate the theoretical maximum throughput of any communication system. The word “information” is used in a precise quantitative sense as a measure of the content of a message, and not in its usual sense to denote semantic meaning [18].

Hartley went on to examine random messages. A random message is defined as a sequence of symbols where each symbol is chosen independently and with equal probability from the alphabet of symbols without regard to which ones preceded it. Since symbols in a random message are equiprobable, there is no way to represent it more compactly, and thus a random message is said to have maximum information content. Its information content in bits per symbol is given by:

$$(H_N)_{max} = \log_2 N, \quad (4)$$

where N is the number of symbols in the alphabet. The maximum rate of information transmission of a communication channel, in bits per second, is:

$$(H_T)_{max} = n \log_2 N, \quad (5)$$

where n the number of symbols transmitted per second. If a message is of length T seconds, the total amount of information transmitted is:

$$H_{max} = nT \log_2 N \text{ bits.} \quad (6)$$

Of course, very few messages are completely random in their structure. For example, in printed English it is clear that letters such as “e”, “t”, and “a” are much more common than “z”, “x”, and “q”. A probability p_i can be assigned to each of the N letters (that is, symbols) in the alphabet.

The sum of these probabilities must, of course, equal 1. Shannon [19] found that the information content of such a message, in bits per symbol, could be determined by:

$$H_N = - \sum_{i=1}^N p_i \log_2 p_i. \quad (7)$$

This equation gives what is called the “0-order information content” since it is based on probabilities that are assigned to the letters without regard to their predecessors. When examining the statistics of printed English, it is obvious that the probability of a particular letter occurring is conditioned by the letter or letters that came before it. For example, the probability that the letter “u” follows “q” is very high, but the probability that “x” follows it is extremely low. When the probabilities are examined in the context of two letters (a “digram”), the analysis is said to be of order 1. Sequences of three letters (“trigrams”), four letters (“tetragrams”), and longer can also be used. Such “N-gram” analyses have an order of $N - 1$.

Shannon [20] found that the information content over N adjacent letters (sometimes called the N-gram entropy) is given by:

$$F_N = - \sum_{i,j} p(b_i, j) \log_2 p_{b_i}(j), \quad (8)$$

or, equivalently:

$$F_N = - \sum_{i,j} p(b_i, j) \log_2 p(b_i, j) + \sum_i p(b_i) \log_2 p(b_i), \quad (9)$$

where b_i is a block of $N - 1$ letters, j is the letter following b_i , $p(b_i, j)$ is the probability of N-gram $b_i j$; and $p_{b_i}(j)$ is the conditional probability of letter j after the block b_i , which is given by $p(b_i, j)/p(b_i)$. The longer the sequences of letters used, the closer these estimates will approach the “true” information content of the message. This is expressed mathematically as:

$$H = \lim_{N \rightarrow \infty} F_N. \quad (10)$$

Entropy of printed English

Plain English was among the first communication systems to be viewed in terms of information theory. The printed language has a fairly standard set of symbols, and these are structured into messages according to well-defined conventions at a variety of levels. Researchers were interested in the efficient transmission of text through a communication system, and thus sought to develop

effective techniques for text compression. From this arose a desire to determine the theoretical limits of compressibility, or, conversely, to find the redundancy of printed text and of natural language in general.

The term *entropy* is used to describe the information content of a message. In connection with printed English, Shannon defines entropy as:

... a statistical parameter which measures, in a certain sense, how much information is produced on the average for each letter of a text in the language. If the language is translated into binary digits (0 or 1) in the most efficient way, the entropy H is the average number of binary digits required per letter of the original language [21].

Redundancy, as the name implies, measures how much can be discarded without losing essential information, in other words "... the amount of constraint imposed on a text in the language due to its statistical structure" [22]. In mathematical terms, the percentage redundancy R of a message is given by:

$$R = \frac{H_{max} - H}{H_{max}} \times 100\%, \quad (11)$$

where H is the measured entropy of the message and H_{max} is the maximum possible entropy of a message in the same system.

The first estimates of the entropy of printed English were made by examining the statistics of representative text, through its N-gram frequency tables. Shannon [23] gives these results for 26- and 27- letter alphabets (the latter including the space character):

	F_0	F_1	F_2	F_3	F_{word}
26 letters	4.70	4.14	3.56	3.3	2.62
27 letters	4.76	4.03	3.32	3.1	2.14

Shannon did not analyze a particular text but extrapolated from published monogram, digram, trigram, and word frequency tables to derive N-gram entropies. His entropy values should be considered as approximations only—especially for units larger than the trigram.

In fact, N-gram analysis is not a viable technique for obtaining an accurate estimate of the "true" information content of natural language. The reason for this is as follows. As N becomes greater

(i.e. the context for the N-gram analysis grows), the entropy estimate will converge to an asymptote (the “true” entropy of the text). However, the analysis is necessarily based on a finite corpus. As N grows, at some point all N-grams will become unique, which implies that the entropy is 0. In principle, this leads to the conclusion that the text is entirely predictable and contains no information at all! This clearly is not the case. For N-gram analysis to work reliably, then, one needs to analyze a huge corpus of text. For example, the Brown corpus [24] contains 1.6 million characters, from an alphabet of 94 different symbols. A trigram analysis is feasible, because 94^3 is less than the size of the corpus. However, a tetragram analysis is not statistically valid since 94^4 is approximately 78 million, far greater than the number of characters in the text.

List of Figures

Figure 1 Using the CHORALE CASINO program to guess notes

Figure 2 Using the CHORALE CASINO program to enter bets

Figure 3 Chorale 151

(a) Cumulative entropy profile

(b) Instantaneous entropy profile

(c) Melody

Figure 4 Chorale 61

(a) Cumulative entropy profile

(b) Instantaneous entropy profile

(c) Melody

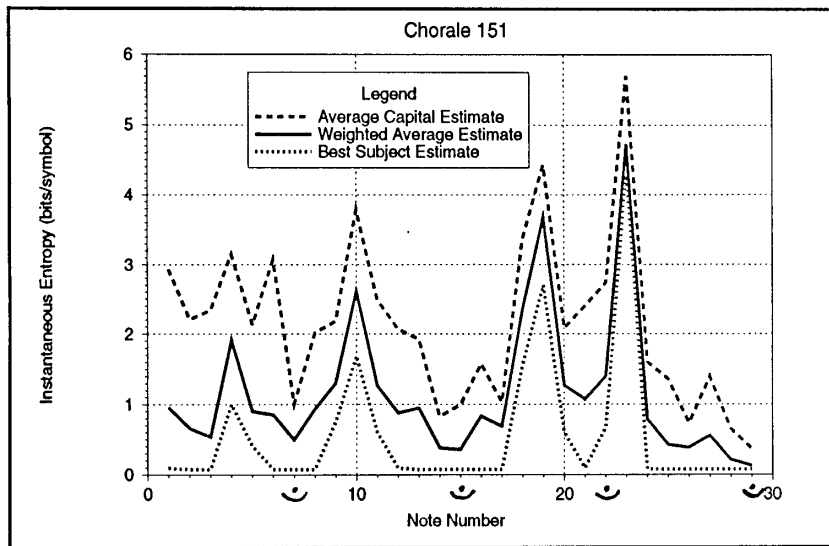
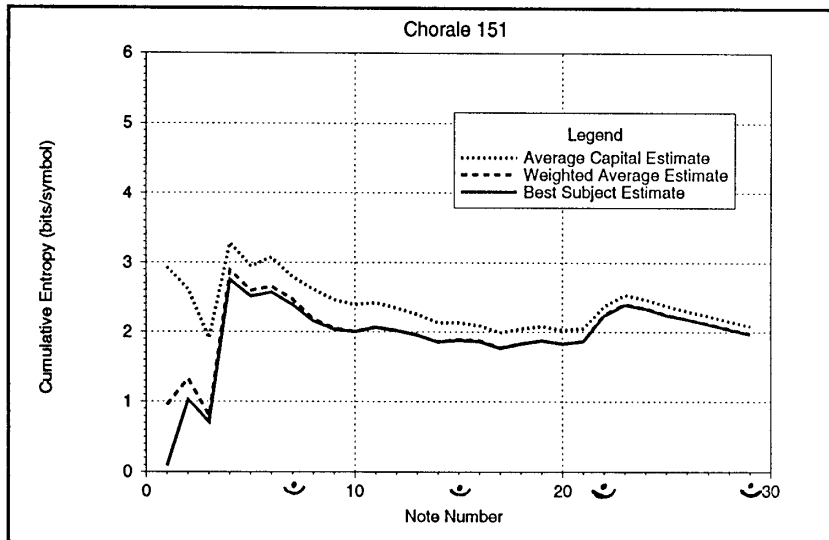


Figure 3

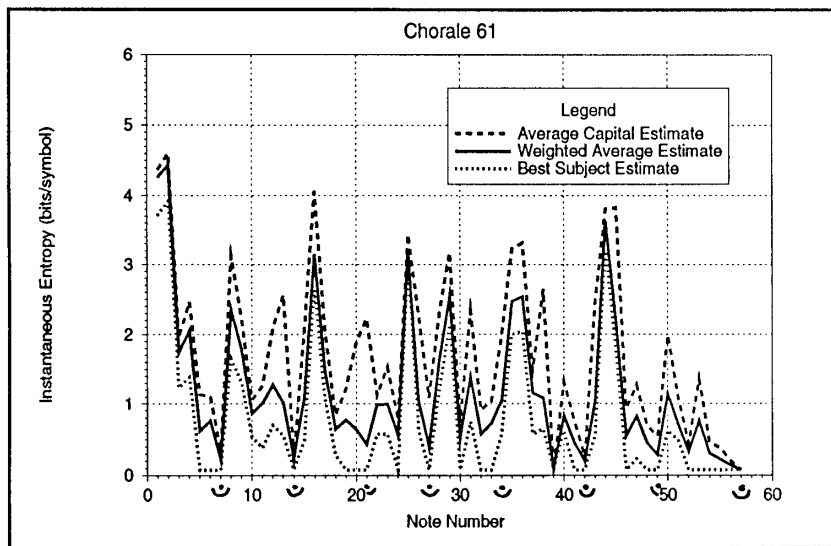
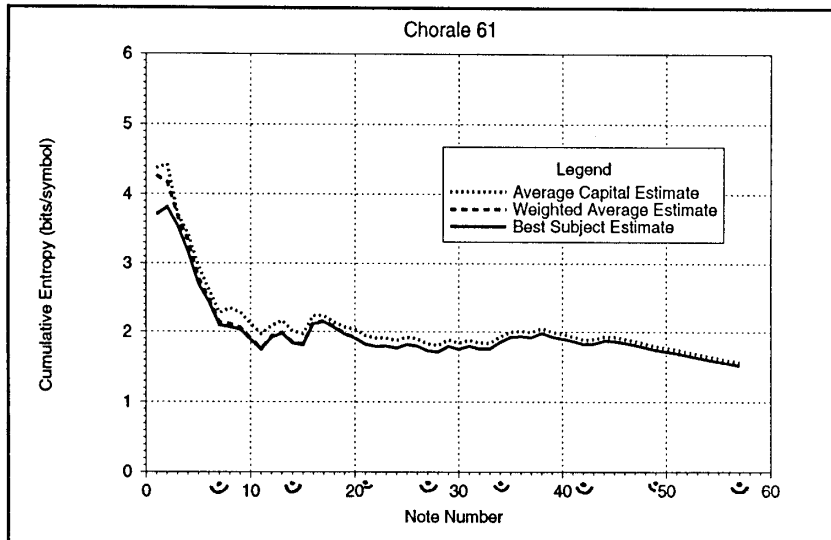


Figure 4