COMPARING ARTIFICIAL INTELLIGENCE MUSIC AND HUMAN MUSIC: A CASE STUDY OF PROF. DAVID COPE'S EMMY SYSTEM

Iris Yuping Ren
University of Rochester

ABSTRACT

This is a pilot study analysing the difference between algorithmic compositions and human music, and subsequently evaluating the authenticity of music-composing algorithms. In this project, we use tools from information theory to study the music pieces created by a well-known system, "Emmy", built by Prof. David Cope. We make the comparison between Emmy's compositions in the style of J.S. Bach's chorales [1] and actual chorales pieces written by Bach [2]. The main metric we use is the Shannon entropy (shortened as entropy in the paper), which measures the unpredictability of information content. We found that the overall entropy of Emmy's pieces are in general higher than the real Bach's pieces, indicating that the AI music are less predictable and are short of repeated notes. We also design the tool of entropy profile and the entropy window to examine the evolution of the entropy within the piece. It turns out that Emmy's pieces have higher and smoother pitch entropy profiles, relatively authentic duration entropy profile, which gives clues on how to improve the composition algorithm: reduce the introduction of new pitch patterns. Taking both the pitch and duration into consideration, Emmy's music have higher entropy profiles throughout, which confirms the observations and conclusions from the whole-piece entropy perspective.

1. INTRODUCTION

Artificial intelligence has been rather successful in many aspects, including creating music. Although has its limitations, there are computer programs that can write new music. According to [3], the time has come in which it is hard for non-musicologists to differentiate between those AI music and the human music. Perhaps there will be one day, even our best musical-minds cannot distinguish the two straightforwardly, if it is not already the case. However, at current stage, we found one natural classifier for distinguishing between human music and human-like AI music generated from the Emmy system: the Shannon entropy from information theory. We will show that there are significant Shannon entropy differences between the two, and this discovery might give insights to how we can improve the existing algorithms.

There has been little done in area of research. Although the idea of algorithmic composition has been there for a long time, Mozart's"Musikalisches Wrfelspiel" (musical dice game) for example, it was not possible to massively and systematically produce these pieces until the recent rise of computer and, subsequently, the computational power. Especially for those algorithms which are imitating a specific genre, it usually requires a training process on a large data set. Possibly because of these recent developments and the subjective nature of the problem, the evaluation and classification of AI composed pieces beyond bare listening and music instinct has not come up. In addition, although the combination of information theory and music goes back a long time, such as the one in [6], it is not completely obvious that the Shannon entropy could be useful in differentiating the AI music and human music. Nevertheless, since Shannon entropy is essentially summarising the probability distribution of event occurrences, we can naturally use this to look at the problem globally and locally. More detailed discussion will be given in the following section.

The organisation of the paper is as follows: In section 2, we will give more attention to the field of AI music. In the next one, section 3, we will give a brief introduction to the tools we used from information theory. Section 4 is the main section to show our calculation followed by discussion and explanation of these results. In the last two section, we will conclude the paper by emphasising our conclusions and discuss future continuation of this work.

2. ALGORITHMIC COMPOSITION

As we mentioned in the introduction, algorithmic composition can found its history back to Mozart times. More recently, famous composers such as Arnold Schoenberg, Anton Webern, John Cage all have composition based on different algorithms. Academically, there has been many systems designed to simulate the music creativity via mathematical models, knowledge-based systems, evolutionary methods, etc. And those methods have been applied to different genres and purposes of music creation: classical, jazzy, improvisational, modernism, atonal music, etc. In this pilot study, we chose a very specific style: J.S.Bach's chorales.

For the AI music data, we got Emmy's music from [1]. The creator of this system, Prof. David Cope, placed 5000 MIDI files of computer-created Bach-style chorales for download, which was very convenient for our experiment. For the real Bach's chorales, we got data from [2].

While Prof. David Cope's book [4] has been disputed a lot, specially in [7], but if we listen to the music he put online, they are actually very close to the Bach's music. Although, even on this similarity topic, people have discussed about the lack of"the meaning of music" when it is written by a computer, we will only consider the subjective metrics instead of the philosophy of AI music. In summary, the Emmy system is well-known in the community of algorithmic composition and has been there for a history of years, so this is not a bad place to start with.

The resource code of Emmy is available online but not well documented. So we will not dive into the details of the algorithm but just analyse the pieces the algorithms have created.

3. INFORMATION THEORY

Information theory has been one of the most important theory in many areas of research. We will talk about the classical Shannon entropy in 3.1 and new dynamic tools we made based on this notion in 3.2. Since we do not have the room here to introduce the whole field of information theory, for more details, please refer to [5], which gives a comprehensive and developed discussion on the topic.

3.1 Entropy

Entropy has the origin of measuring the storage needed to communicate a message. Mathematically, it is defined as in Eqn (1). It is the expectation value of the information content, which is defined as $-logp(x_i)$.

$$H(X) = -\sum_i p(x_i)logp(x_i), i \in n = outcomes \quad (1)$$

Intuitively speaking, we get more information from an unlikely event, and therefore an event with low probability give us more information. For example, if a string is 11111..., from the distribution $P(1) = 1$, we do not get lots of information from the next number, which is bound to be 1, simply because there is nothing changing and we do not need much space for storing this message; the string 15342..., from the distribution $P(x) = 1/n, x \in 1, 2, 3..., n$, on the other hand, is something we do not know its pattern and not be able predict the next occurrences, and therefore gave us more information with respect to the knowledge we already know. Another interpretation from this is that, a high entropy indicates a surprisal element (get to maximum when the input follows uniform distribution), while a low entropy indicated a more predictable pattern (get to minimum when the input follows constant distribution).

In the case of music, we can calculate the entropy of a piece by counting the frequency of musical events. For example, we can count the appearances of each note/ pitch/ duration and get the discrete distribution of those musical events in the piece. Then we can use the equation to calculate the information content of each note and then take the expectation to obtain the entropy in the end. So, essentially, the entropy is tied with the frequency of musical events in a specific range. This process is taking out of the consideration of the events' order. As we are smearing out the time dimension, the different behaviors of entropy stem from the differences of how many different types of musical events there are and how repetitive they are.

To introduce more properties on the relationship between entropy and repeating patterns, we introduce some properties as follows:

- For uniform distributions, the entropy increase with the number of outcomes. For example, the entropy of a monotone sequence is lower than the entropy of scales of any kind.
- The entropy is lower when there is reduced uncertainty. For example, tonal music has smaller entropy than atonal music, since there are more frequent note in the tonal, and this give us a smaller term in the definition of entropy.
- The entropy remains the same when there is a repeat. For example, if we concatenate 2 bars of the same music together, the 4 bars music will have the same entropy as the 2 bars music.

The proofs of those properties are trivial from the definition of the entropy.

3.2 Entropy profile and entropy window

The above definition of entropy is a good measure from a global point of view. However, we lose lots of local information because of the summation in the end and assuming we know the whole distribution from the beginning. We therefore propose two extended tools based on entropy: the entropy profile and the entropy window.

We calculate the entropy window by taking chunks of equal length from the music and calculate the entropy on the chunks. Formally, let l be the window size, N be the length of the note sequence in a piece, the entropy window is a vector E_i of length Nl, where E_i is the entropy on window $n_{i,i+l}$, can be expressed in equation 2, in which $X_{k_i} = \{X_i, X_{i+1}, ..., X_{i+l}\}$.

$$EW(X) = \{H(X_{k_1}), H(X_{k_2}), ..., H(X_{k_n})\} \quad (2)$$

Basically, we localise ourselves to the just one window of music to calculate the entropy just as the way it was before. Finally, we take the vector generated by the chunks' entropy, and call it the entropy window vector.

For the entropy profile, similarly, we take chunks of music and calculate the entropy, but instead of just windowing, we take chunks of increasing length, with the same increment. It is equivalent to adding up all the entropy window before a certain time, that is, to retain memory in music. Formally, let l be the window size, N be the length of the note sequence in a piece, the entropy window is a vector EP_i of length Nl, where EP_i is the entropy on window $n_{1,i+l}$, can be expressed in equation 3, in which $X_{m_i} = \{X_1, X_2, ..., X_{i*l}\}$.

$$EW(X) = \{H(X_{m_1}), H(X_{m_2}), ..., H(X_{m_n})\} \quad (3)$$

By looking at the entropy profile and entropy window, we can understand more on the evolution of music entropy through time. The difference between the two is that, we

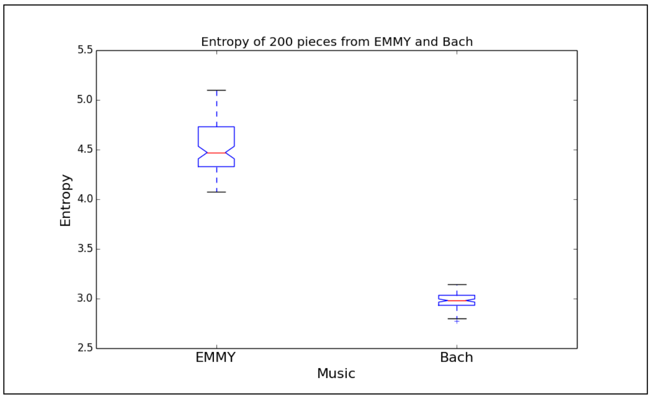

Figure 1. Entropy of pitch and duration pair. X-axis corresponds to EMMY's music and actual Bach's chorales. The AI music and human music are separate: the entropy of music is a natural classifier of Emmy's chorales and Bach's chorales.

take into consideration of history when calculating the entropy profile, while we are considering no memory effects when calculating the entropy window. The reality should actually be in-between: we memorise excerpts of music but not all of them when we listen. And one more deviation from the reality is that, we do not consider prior musical experiences, which is usually not true. Nevertheless, this study is not trying to mimic entirely the cognitive process of music listening, but to use generic tools to understand better on the differences between AI music and human music.

4. RESULTS

In this section, we will show our results of the entropy, entropy window and entropy profile calculations. Although we have performed the calculations in hundreds of pieces, it is not possible to show them all here. Also, because there are different length of music, we are taking the minimum length of all the pieces in the comparing sets. The values for the whole pieces were calculated, they follows the same trend as the music in the minimum length. More figures can be provided if requested.

4.1 Entropy

As shown in Figure 1, used the definition introduced in 3.1, we can observe immediately the huge entropy difference between the AI music and human music. It is almost trivial to build a classifier between the two based on this result. This suggests that, although for hard for non-trained ear to distinguish this similar-to-human music, there are underlying intrinsic properties the Emmy system is not fulfilling, and can manifest themselves via the entropy calculation. This provides a convenient way to differentiate out the AI music by using pure computations on note frequencies while not incorporating any musical knowledge.

We will see more elements behind this phenomenon in the next subsections.

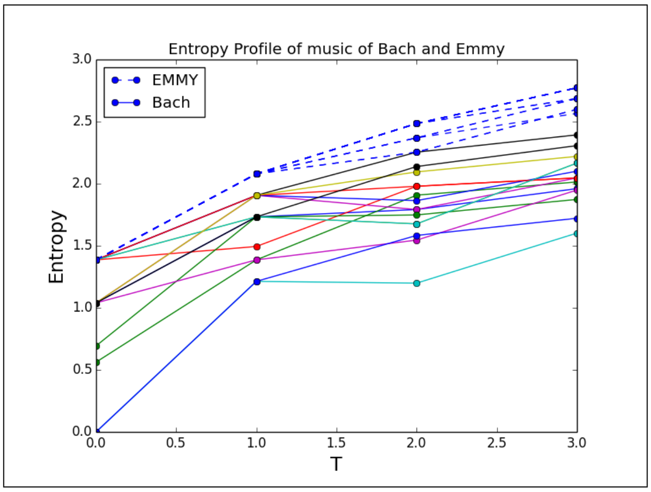

Figure 2. Entropy profile of pitch and duration pairs. The direction of x-axis is the time evolution in music, and the unit of x-axis is the window length, which is 4 notes in this figure. The y-axis indicates the entropy. The dotted lines are entropy profile from AI music and the solid lines are from human music. The AI music and human music have different patterns. The unnaturalness is showing itself in the growing smooth curves.

4.2 Entropy profile

In this subsection, we examine the entropy discrepancy between the AI music and human music in a more detailed way. First, as introduced in 3.2, we use the entropy profile to see the temporal evolution of the music entropy. As shown in Figure 2, we observe that actually most of the AI music pieces starts already from a larger entropy value, and continues to grow, seldomly decrease, while the human music shows slower increase trend and decrease more often. Using the properties introduction in 3.1, we know that some possibilities for this are: there is not a prominent enough note in the AI music, and more generally, there is a lack of repeated pattern in the music. Cognitively speaking, this corresponds to the fact that, when listening to music, we form the correct music pattern expectations after hearing the music for a while, there should be no more surprisal, but from time to time, the music surprises us again. Indeed, it seems to be a common problem that the computer programs are sometimes "too innovative" in terms of writing music, and not giving us enough "musically interesting surprisals".

Furthermore, we look at the entropy profile from a pitch-only and a duration-only perspective to try to see if the problem lies in one particular side , as shown in Figure 3 and Figure 4. In accordance with our hypothesis, we found that the problem lies more with the pitch entropy profiles. The duration entropy profiles are quite well mixed together, while the AI pitch entropy profiles grows in a even faster and smoother way. The smoothness must have come from the Emmy's algorithm itself, to which we do not know much. However, this is surely creating some gaps between the AI music and the human music. So, as we talked in the last subsection, one possible way to disrupt the smooth pat-

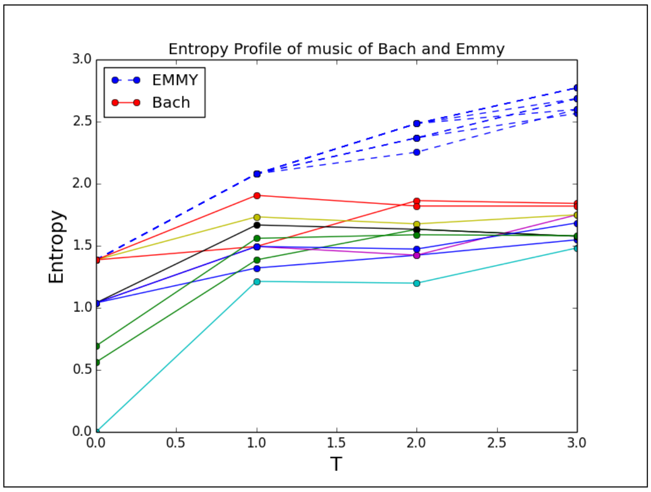

Figure 3. Entropy profile of pitch only. The specifics of the figure are the same with Figure 2. The AI music and human music have different patterns. The unnaturalness is showing itself in the growing smooth curves.

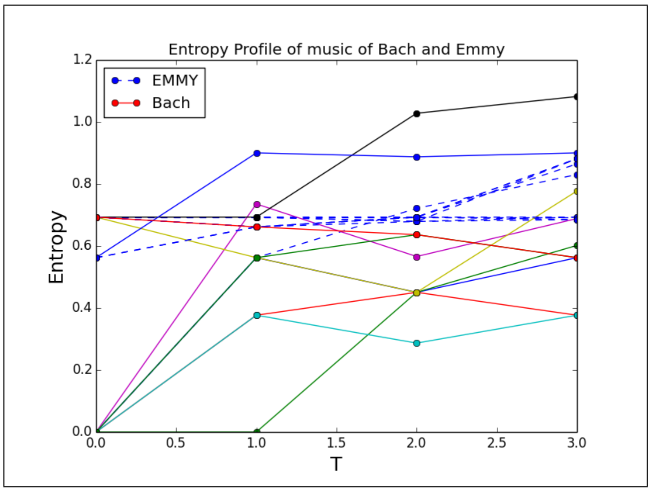

Figure 4. Entropy profile of duration. The specifics of the figure are the same with Figure 2. The AI music and human music are relatively well-mixed.

tern is to introduce more repeated pitch patterns and prominent pitches. In general, the alternations between different trends of entropy is needed to be introduced into future composition algorithms to improve, at least at the very first level: pitch frequency.

4.3 Entropy window

Took off the memory of the system, we will be looking at the entropy window. To summerise the windowed entropy of more pieces, we first plotted the boxplot of the pitch and duration pairs from 40 Emmy's pieces and 40 Bach's pieces, as shown in 5. The rest are similar with the structure of the last subsection, we shows the entropy window introduced in 3.2 for pitch only and duration only in Figure 6 and Figure 7, respectively. We can see that the local properties of entropy without long memory of the piece is still problematic, especially pitch-wise. The duration, on

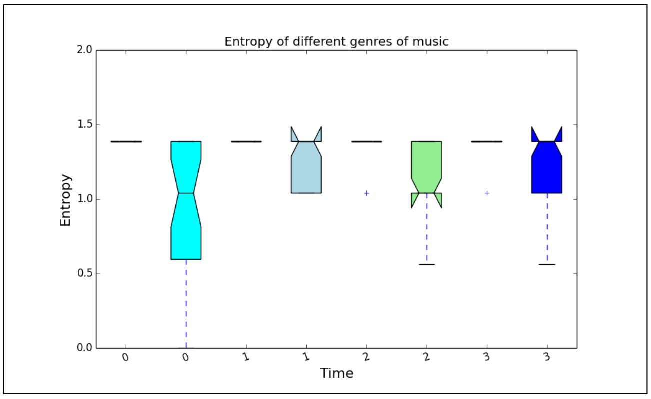

Figure 5. Entropy window boxplot of the pitch and duration pairs of 40 Emmy's and Bach's pieces. The direction of x-axis is the time evolution in music, and the unit of x-axis is the window length, which is 4 notes in this figure. There are always two of the same x values because each of them represents the Emmy's music and Bach's music respectively. The AI music and human music have different patterns, with the AI music are always reaching their maximum entropy.

the other hand, is relatively well mixed and less problematic.

The AI pitch entropy window value is basically reaching the maximum entropy (follows uniform distribution, that is, the entropy calculation on vector $(0.25, 0.25, 0.25, 0.25)$, gives $4 * (0.25 * log_e 0.25) \approx 1.38)$ in the pitch and pair entropy window figures and staying put. Notice, on the contrary, the actual Bach's pieces duration entropy window sometimes reaches 0, and this is because Bach is using consecutively 4 same durations in this window, thus, $4 * (1 * log_e 1) = 0$. Just to see if this is true throughout the whole pieces, in Figure 8, we are showing the entropy window of the whole piece without trimming the AI music length to the Bach's pieces length. We can see that there are only a few times that this entropy value drops. This means, similar to the discussion on entropy profile, given most of the excerpts of these AI pieces, there is room for improvement just by adjusting the pitch frequency within each bar: make some notes/patterns more frequent than others, and therefore reduce the entropy.

5. SUMMARY

In summary, we calculated entropy, entropy profile and entropy window on the human-like-AI music (Emmy's chorales in Bach's style) and human music(J.S.Bach's chorales), and found big differences between the two. This indicates that the information measure of entropy is an important property in algorithmic composition, and further algorithm could verify and improve their system by calculating these values when generating music. For example, when the entropy is too high, we can reduce introducing new musical events and including more repeated patterns and prominent notes. We do not know yet if our results could be extended to other system, but if so, entropy could be an intrinsic quan-

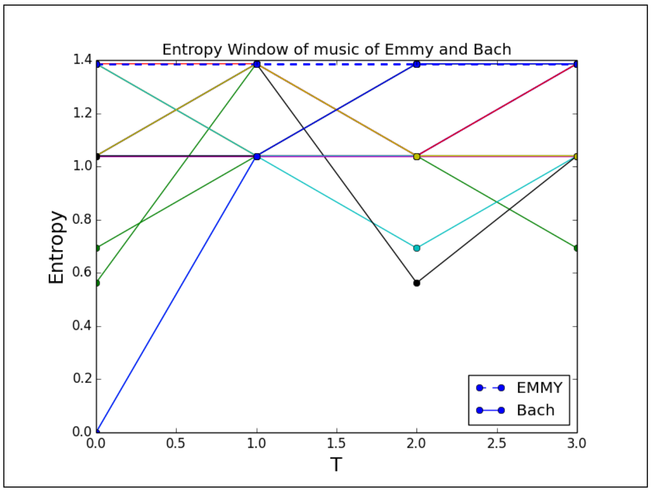

Figure 6. Entropy window of pitch only. The specifics of the figure are the same with Figure 2. The AI music and human music have different patterns, with the AI music are always reaching their maximum.

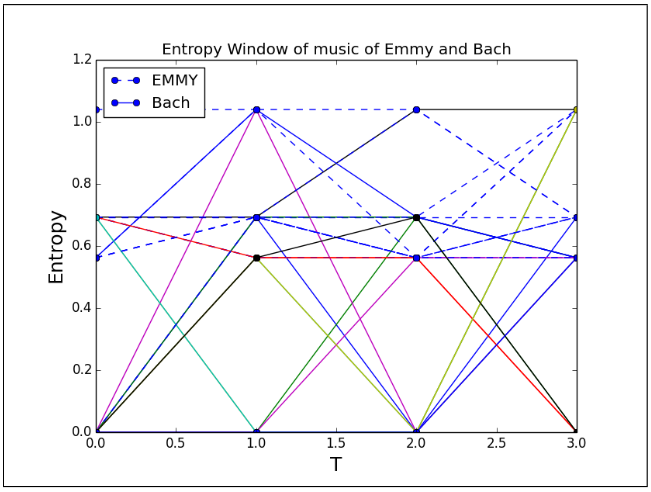

Figure 7. Entropy window of duration only. The specifics of the figure are the same with Figure 2. The AI music and human music are relatively well-mixed.

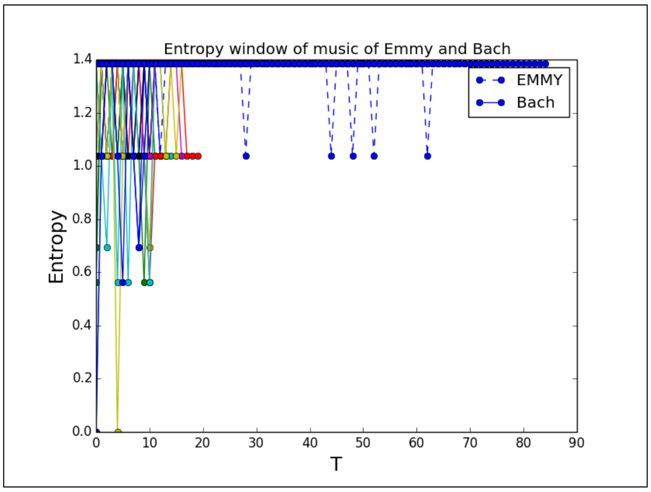

Figure 8. Entropy window of pitch and duration pair. The specifics of the figure are the same with Figure 2. Whole piece displaying. The long dotted line and short solid line is due to the different length of different pieces. We plotted only the first few windows and profile time in the previous figures also because of this.

tity to distinguish human-like-AI music and human music. Otherwise, we can regard the entropy as a guideline to improve composition algorithms and make the gap between human-like-AI music and human music smaller.

6. FUTURE WORKS

As the field of AI music develops, we believe the tools from information theory can be more and more helpful in evaluating and improving music. The first step of further developing would be to verify that the conclusion in this study can be applied to more algorithmic composition system. For that purpose, we would need more data from different such systems and corresponding comparable human music, or develop and realise some toy examples and theoretical systems to test out more closely, and try to find if there is a system can generate music which have the same entropy properties as the human music.

In the meantime, there is also space for the information measure to be further improved: instead of just taking the notes probability distribution from each piece, to further mimicking the cognitive process of a real music listener, we should better have an assumption of the underlying distribution of the listens gained from their musical exposures. One more fundamental problem with the application of entropy is that we do not have the order of the notes considered, since entropy is defined only via distribution of occurrences, therefore ignoring which notes comes first and which comes after. To incorporate the order element, we can perhaps look at the entropy of repeated musical patterns themselves or come up with a relevant new measure.

Moreover, going beyond information theory, we would like to develop a more complete sets of methods to evaluate music-generating algorithms. Most importantly, including the cognition process of people listening to music and evaluating music.

7. REFERENCES

[1] 5000 works in bach style. http://artsites.ucsc.edu/faculty/cope/5000.html.

[2] Bach's chorales. http://kernscores.stanford.edu/.

[3] Undiscovered bach? no, a computer wrote it. http://www.nytimes.com/1997/11/11/science/undiscovered-bach-no-a-computer-wrote-it.html?pagewanted=all.

[4] David Cope. *Computer models of musical creativity*. MIT Press Cambridge, 2005.

[5] David JC MacKay. *Information theory, inference and learning algorithms*. Cambridge university press, 2003.

[6] Leonard C Manzara, Ian H Witten, and Mark James. On the entropy of music: An experiment with bach chorale melodies. *Leonardo Music Journal*, pages 81–88, 1992.

[7] Geraint A Wiggins. Computer models of musical creativity: A review of computer models of musical creativity by david cope. *Literary and Linguistic Computing*, 23(1):109–116, 2008.