Mary: An AI music composer.

Nick Thomas

January 22, 2012

Contents

1 Overall Design **2**
1.1 Notes, Phrases, and Concepts 2
1.2 Fitness Function 3
1.3 Composition Algorithm 5

2 Fitness Function **6**
2.1 Mapping Metrics onto a Normal Distribution 8
2.2 Harmonic Consonance 8
2.3 Melodic Consonance 8
2.4 Rhythmic Consonance 8
2.5 Novelty 8
2.6 Richness 8

3 Concepts **9**
3.1 Concept Language 9
3.2 Analysis and Synthesis 9
3.3 Transformations 9
3.4 Additions 9

4 Composition Algorithm **10**

Chapter 1

Overall Design

1.1 Notes, Phrases, and Concepts

A phrase is a piece of music, of any size. It could be anything from a short melody to a complete composition. A phrase is a collection of notes.

What is a note? It has four properties: pitch, velocity, position, and duration. Notes are arranged in a theoretical piano roll which has 88 pitches (for the 88 keys of the piano keyboard). The time axis of the piano roll is divided into "ticks." Velocity is between 0 and 127, like in MIDI.

Here it is worth saying a few words about Mary's time scheme. A tick is a 16th note. Mary composes without thinking about the tempo of her music. We can play it back at whatever tempo pleases us.

Mary has "concepts." These are patterns which phrases can match. They let Mary think abstractly about music. Hypothetical examples of concepts are "upwards diatonic scale," "minor chord," "syncopated rhythm," "ABA composition scheme," etc.

A phrase records the concepts that it matches. It also records the concepts that its sub-phrases (i.e., portions of the phrase) match. We can think of a phrase as a topological space over a collection of notes, where each subset in the topology matches some concepts.

Concretely, a note n is a quadruple (p, v, s, d), where:

- $p \in [0, 87]$ is the pitch.
- $v \in [0, 127]$ is the velocity.
- $s \in \mathbb{N}$ is the position.
- $d \in \mathbb{N}$ is the duration.

A phrase P is a quadruple (N, C, T, D), where:

- N is a set of notes.

- C is a set of concepts which P matches.
- T is a set of phrases, each of which has notes that are subsets of N. In other words, for all $(N', C', T', D') \in T$, $N' \subset N$.
- D is the dimensions of P.

The dimensions D record the minimum and maximum pitch and position of P, as well as the length and pitch range. D is a sextuple $(s^-, s^+, s^r, p^-, p^+, p^r)$, where:

- s^- is the minimum position. In other words, it is the largest number such that, for all $(p, v, s, d) \in N$, $s^- \leq s$.
- s^+ is the maximum position, defined similarly.
- s^r is the length of the phrase. $s^r = s^+ - s^-$.
- p^- is the minimum pitch, defined similarly to s^-.
- p^+ is the maximum pitch, defined similarly.
- p^r is the pitch range. $p^r = p^+ - p^-$.

We can think of D as a box drawn around the phrase in the hypothetical piano roll, demarcating its borders.

1.2 Fitness Function

To write an AI which makes beautiful music, one needs an algorithmic definition of musical beauty. We need a “fitness function,” which takes a phrase as input and outputs a number saying how beautiful it is.

The task of designing a fitness function, though not the most technically problematic part of Mary, is certainly the most philosophically problematic part. Is it even possible to write a function quantifying musical beauty? And if so, how do we do it?

The question of the possibility of quantifying musical beauty depends on several more general questions. Are the qualia of aesthetic beauty reducible to mathematics? Are the human mental processes which judge music reducible to mathematics? Are human mental processes in general reducible to mathematics? Are things in general reducible to mathematics?

Our question ultimately hinges on the question of whether or not reductionism is true. If reductionism is false, then we would expect it to be impossible to write a function quantifying musical beauty. If reductionism is true, then we would expect it to be possible to write such a function.

Let us suppose that reductionism is true. In this case the mind is the brain, and human aesthetic judgments are reducible to neurological processes. Our fitness function simply has to be something which does something like what those human neurological processes do.

In this case, there are still potential obstacles to writing our fitness function. Maybe the needed function is so complex that we can't manage to figure it out.

Furthermore, we would expect the processes of aesthetic judgment to vary somewhat from person to person. There would not be a single, objectively correct metric of musical beauty; rather, there would be many metrics of musical beauty, corresponding to the different metrics that different human brains implement. All of these metrics would be rather similar to each other. But we could not identify a single function and say, "this function, and only this function, defines musical beauty."

This latter point, while philosophically significant, probably does not make a big difference in practical terms. We will be doing well enough if we can write a fitness function which is comfortably inside the space of human beauty metrics.

Now let us consider the harder case: the case where reductionism is false. In this case we can hope for less. But even in this case, beauty cannot be something so ineffable and irreducible that there is literally nothing we can say about it.

Consider the fact that people like consonant intervals, regular rhythmic patterns, and music which repeats itself. Consider the fact that people like the sounds of violins and TB-303s. Consider, even, the fact that people like a breakdown in the middle of a heavy dance track. All of these facts indicate some quantifiable regularity in the sorts of music that people like.

Maybe we cannot say everything about what is beautiful. But we can, at any rate, say *some* things about what is beautiful. Maybe we can say enough to let a computer write good music. Only one way to find out!

I have designed a fitness function which I think captures some of what we like in music. It uses three metrics: consonance, novelty, and richness.

"Consonance" here has a broader meaning than it does in ordinary usage. In general the term "consonance" refers to the use of consonant harmonic intervals: perfect fifth as opposed to minor second, for instance.

We call this type of consonance "harmonic consonance," and we also define two other kinds: melodic and rhythmic consonance.

Melodic consonance is the extent to which a phrase forms smooth and clearly defined melodic lines. A phrase which moves in small, progressive steps across the scale is melodically consonant; one which randomly jumps around on the scale is melodically dissonant. So, for instance, a motion of major seconds is melodically consonant, while a motion of minor sevenths is melodically dissonant.

Rhythmic consonance is the extent to which the rhythms in a phrase are regular. A 4/4 kick drum is rhythmically consonant; the noise a modem makes when it connects to the Internet is rhythmically dissonant.

People like music which has these types of consonance; but if consonance were our only metric, it would produce very dull music. We would just have C major chords played in a 4/4 pattern for three minutes.

This brings us to our second metric: novelty. Novelty is the extent to which a phrase is distinctive, unique, and surprising. This is also clearly something that people like in music. Think of the fact that a song we have heard many times starts to bore us. Think also of the fact that music changes dramatically over time. If people didn't care about newness in music, we wouldn't expect that to happen.

Consonance and novelty are opposing considerations. It is easy to make surprising, unique music by banging chaotically on a keyboard; but the resulting music is harmonically, melodically, and rhythmically dissonant. On the other hand, a C major chord in a 4/4 pattern is very consonant, but not at all surprising. The need to balance novelty and consonance ought to result in interesting and pleasant music.

These two capture some important aspects of musical beauty. But there are also things that they don't capture. What about the fact that a long piece of music has a thematic unity and regularity in its structure? A piece could be both consonant and novel, but randomly change its mood from happy, to melancholy, to angry, and fail to create any definite impression.

I address some of this problem with a third pair of metrics: unity, and variation. A phrase which repeats itself (that is, has sub-phrases matching the same concepts) has unity. Since we don't want a piece to just repeat the same thing for five minutes, we have an opposing metric: variation. A phrase which diverges from itself (that is, has sub-phrases matching different concepts) has variation. A phrase is supposed to have a balance between unity and variation.

After this, there still seems to be something important that hasn't been said. I think of really stirring pieces of music, and it is hard for me to imagine that their stirring nature stems from these mathematical considerations. It feels like the composer has tapped into something deeper than anything that a calculation could reach, and somehow given it concrete expression. I have the intuition that neither these metrics, nor any conceivable metrics, say what beauty really is.

1.3 Composition Algorithm

Chapter 2

Fitness Function

The fitness function quantifies how good a phrase is. It is intended to approximate the concept of musical beauty. It is based on three metrics: consonance, novelty, and richness.

Consonance has three sub-measures: harmonic consonance, melodic consonance, and rhythmic consonance. Harmonic consonance measures the consonance and dissonance of the chords and intervals in the phrase. Melodic consonance measures the extent to which the phrase forms smooth and distinct melodic lines. Rhythic consonance measures the regularity of the rhythms in the phrase.

Novelty is a measure of how unique, surprising, or distinctive the phrase is, relative to the phrases that Mary usually sees. It is measured statistically.

Richness is a measure of the conceptual complexity of the phrase. Phrases which match more concepts are richer. Richness also has two sub-measures, having to do with the conceptual complexity of the phrase's sub-phrases: unity, and variation.

When sub-phrases match the same concepts as each other, this increases the unity score. When sub-phrases match different concepts from each other, this increases the variation score. A balance between unity and variation is considered desirable.

Each metric is given a score from 0 (maximum consonance) to 1 (maximum dissonance). These scores are computed statistically such that the phrases that Mary sees follow a normal distribution on all metrics.

Each metric is squared while combining them, to capture the idea that it is more important that a phrase have a little of every metric than that it have a lot of one metric. This encourages Mary to look for phrases that have a balance between the metrics.

There are also configurable coefficients which weight the different metrics relative to each other.

Let us prepare to write the formula for fitness. Let variables be defined as in Table 2.1.

F	Fitness
c	Consonance
C	Consonance weighting coefficient
h	Harmonic consonance
H	Harmonic consonance weighting coefficient
m	Melodic consonance
M	Melodic consonance weighting coefficient
y	Rhythmic consonance
Y	Rhythmic consonance weighting coefficient
n	Novelty
N	Novelty weighting coefficient
r	Richness
R	Richness weighting coefficient

Table 2.1: Variables in the fitness formula.

The coefficients for the three metrics must add up to one. The same is true of the coefficients for the three sub-metrics of consonance. In other words, let:

$$C + N + R = 1 \tag{2.1}$$

$$H + M + Y = 1 \tag{2.2}$$

The formula for consonance is:

$$c = \sqrt{Hh^2 + Mm^2 + Yy^2}. \tag{2.3}$$

The formula for fitness is:

$$F = \sqrt{Cc^2 + Nn^2 + Rr^2}, \tag{2.4}$$

which is equivalent to:

$$F = \sqrt{C(Hh^2 + Mm^2 + Yy^2) + Nn^2 + Rr^2}. \tag{2.5}$$

I conjecture that F will follow a normal distribution. I have yet to verify or falsify this.

2.1 Mapping Metrics onto a Normal Distribution

2.2 Harmonic Consonance

2.3 Melodic Consonance

2.4 Rhythmic Consonance

2.5 Novelty

2.6 Richness

Chapter 3

Concepts

3.1 Concept Language

3.2 Analysis and Synthesis

3.3 Transformations

3.4 Additions

Chapter 4

Composition Algorithm