CHAPTER ONE

Introduction

Whence Musicat?

In 2004, I came to graduate school at Indiana University in Bloomington with a very specific research goal: I wanted to make a computer program that could write music. More specifically, I intended to write a program that could compose its own beautiful and original melodies[1] — not algorithmic music in the tradition of Lejaren Hiller and Leonard Isaacson's *Illiac Suite*, or "recombinant" music such as the Bach-derived inventions and Joplin-derived rags of David Cope's EMI. I had a specific plan: I would create a composing program modeled after ideas of the program *Letter Spirit*, from the Fluid Analogies Research Group (FARG), which was intended to carry out the creative task of generating stylistically self-consistent alphabets. Specifically, I was excited by a list of bullet points on page 411 in the book *Fluid Concepts and Creative Analogies* (Hofstadter & FARG, 1995) that gives a set of requirements that must be met for a program to be called "creative". However, after arriving in Bloomington, I quickly realized that it was premature to attempt the computer music-composition endeavor, because of a problem that became more apparent to me as a student

[1] This imaginary melody-creation machine, like George Orwell's melody-composing "Versificator" in the novel *1984*, is just the sort of thing that Douglas Hofstadter, my advisor, doesn't like the idea of, and doesn't expect will happen any time soon. See (Cope & Hofstadter, 2001) for a lively discussion.

of cognitive science than it had been earlier when I was simply excited about the prospect of computers generating music: *computer programs can't listen*. My imagined computer composer would theoretically come into existence with the ability to *create* melodies yet without any ability to *hear* them. This made no sense, since I couldn't expect high-quality musical results to issue from a deaf computer. (Beethoven composed music while deaf, but he had had the experience of many decades of listening to and composing music while he could still hear.) Therefore, I refocused on an activity that is fundamental to the task of music composition: music *listening*.

What is Musicat?

This thesis thus describes my first efforts at computationally modeling human music-listening, implemented in my computer program called "Musicat". (The name "Musicat" was inspired by the program from which the core architecture of Musicat is derived, Copycat, along with Copycat's successors, Metacat and Magnificat.) The program simulates the process of listening to a simple melody in the Western tonal tradition, and displays on the screen the various internal cognitive structures that form as the melody progresses in time. These structures take the form of *groups of notes* and *analogies between group structures*. For example, Figure 1 shows the so-called "Fate" motif at the start of Beethoven's Fifth Symphony. What mental representations are formed when a *person* hears these notes?

Figure 1.1: Opening motif of Beethoven's Fifth Symphony.

Several things might happen for a listener. First, after the first note is repeated, and then repeated again, any listener will notice that there are three instances of the same note in a row; these three notes might be mentally grouped together. Then, after the fourth note sounds and is held for a while, all four notes will likely be grouped together. This kind of grouping is predicted by theories of gestalt perception. Much more cognitive activity is possible, of course. The listener will notice the different fourth pitch, and will probably recognize it as downwards motion, or even motion by a certain distance (a leap of a "third"). A listener familiar with British war announcements broadcast via radio in World War II might feel a variety of complex emotions if reminded of the use of these four notes in that particular context.

Figure 1.2: Opening motif, including the next four notes.

After the fermata, things get a bit more interesting (Figure 1.2). Several of the same processes occur when a person hears these next four notes: the first three notes may be heard as being three instances of the same pitch, they may be grouped together, the following note will sound like a leap downwards, and so on. But the pitches are different; this new passage sounds like a version of the first passage that has been shifted down in pitch. After the music reaches the second fermata, the listener clearly will have heard two main groups of notes: the first four notes, grouped together, followed by the next four notes, also grouped. Not only has the listener formed a mental boundary between the two groups, but also a connection between them has been established. In short, an analogy is perceived between the groups (Figure 1.3).

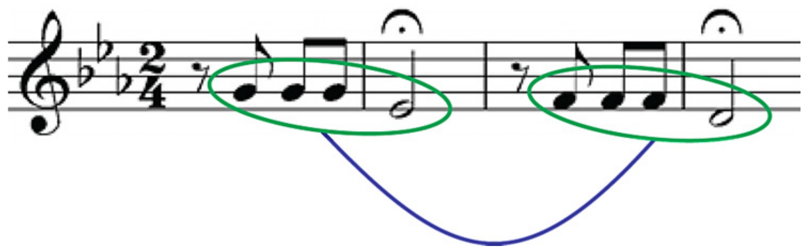

Figure 1.3: Previous figure with groups and an analogy.

We see that there are two groups of notes, surrounded by ellipses, and there is an analogy between the groups, indicated by an arc. This is the type of mental representation modeled by Musicat. The program's goal is to *simulate real-time listening by creating internal representations (such as groups and analogies) of the musical structures it "hears" as music is given to the program one note at a time*.

Wherefore Musicat?

But why, one may ask, is modeling music-listening an interesting problem? The example above looks very simple. And isn't listening a passive activity? People listen to music all the time, without apparent effort. These days, computers are supposed to be great at recognizing patterns, as well — this is the era of statistical machine learning and "Big Data", after all! — so can't we just run an off-the-shelf "pattern recognizer" on music and be done? Smartphones can already "listen" to and identify songs on the radio. So, what is it about how *people* listen to music that makes it worthy of study?

I argue that listening is a much more active process than many people give it credit for. There is much more to listening than meets the ear, as I will now explain.

Music-listening as Performance

Listening might seem so passive that one might expect it to be simple (or pointless) to model. What would be the point of a computer that didn't do anything but sit on a desk and listen to songs? I will make three claims about listening to show that it is a more active process than most of us probably realize:

1. Music-listening is dynamic.
2. Music-listening is a creative process.
3. Music-listening is not a universal talent.

In order to emphasize that music-listening involves cognitive work, I introduce the term "listening performance" to refer to the *creative* act of listening.

Music-listening is Dynamic

> "To stop the flow of music would be like the stopping of time itself, incredible and inconceivable." — Aaron Copland (Copland, 1960)

Music-listening is a temporal process. Whereas we can see an entire painting at one time, music comes to us as a *series* of notes, with new notes coming into our conscious awareness as older ones quickly fade away. Just this temporal aspect alone might make one think that music requires focused attention in listening. Music is even more dynamic than this suggests, however.

One might at first be tempted to think of serious, focused listening as follows: it is a mental process that registers each new note or chord as it is sounded, dutifully paying attention to the passing music in the same way that a copy machine or scanner passes a light over a piece of paper and transfers the original to the copy, one horizontal slice at a time. After all, this is how musicians perform music — one note after another, until a piece is finished. Conductors or drummers beat out musical time in a 1–2–3–4 pattern that repeats over and over, always moving forward. Time flows in only one direction. Surprisingly, however, this portrayal is misguided, because listening is temporally flexible: it can be *retroactive*. An example follows (see the "Mystery Melody" in Chapter 3 for another example).

Example: Ants Marching

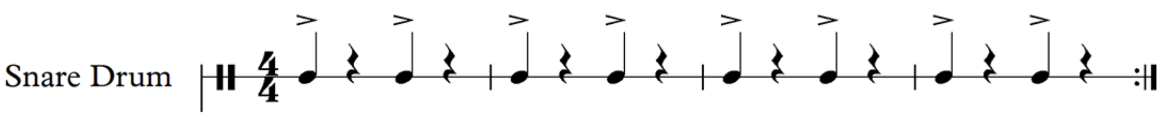

Figure 1.4: Opening of "Ants Marching".

Consider the simple introduction to the song "Ants Marching", by the *Dave Matthews Band* (Figure 4). The introduction varies in length in different live performances, but always features an extremely simple drum pattern (a single, loud snare-drum attack) repeated over and over. In one recording, this continues for over a minute, driving the percussive sound into the listener's ear almost *ad nauseam*. But eventually, a big surprise occurs: the band enters on an offbeat (Figure 1.5), playing a loud chord with a prominent falling tenor saxophone sound accentuating the first chord:

Figure 1.5: Opening melodic motif in "Ants Marching".

The drumbeat happens again, and in the next measure, there are three more chords in the fiddle, with a simple melody on notes F♯ and G in the lower voice. At about this point for me, mental confusion ensues and I feel disoriented as I reinterpret the metrical structure. A moment later, the rhythmic structure suddenly falls back into place for me and I re-hear the music in the way notated in Figure 1.6, with the drum pattern shifted to the *offbeats*:

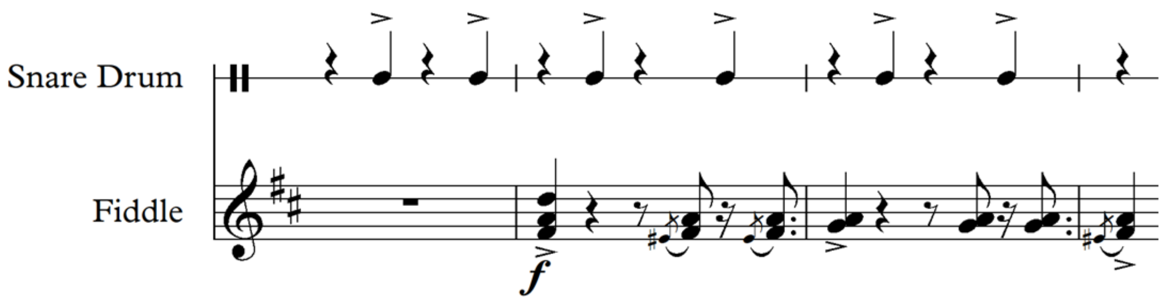

Figure 1.6: Rhythmic reinterpretation of Figure 1.5: drum on beats 2 and 4, not 1 and 3.

Intellectually, I now realize I had been "tricked" the whole time by the drum, which started the song a minute earlier by playing on an offbeat, not the downbeat, as I had assumed. It feels, however, that something "shifted" in a hazy, disorienting way as the first notes of the actual melody were playing. Specifically, I had to reinterpret and rehear several drum attacks, even including a few that occurred before the fiddle's entrance, as falling on beats 2 and 4, not 1 and 3.

This last phenomenon is what I mean when I claim that music-listening is a dynamic and retroactive process. Mental interpretations of musical structures are not static, and may even change retroactively, after they have been formed. Although we can't go back in time, we do hold recently-perceived music in working memory for a brief time, or the order of seconds (perhaps by using a mechanism known as the *phonological loop* that allows several seconds of music to be heard in the "perceptual present"— see Chapter 2), and we can indeed go back and revise our perceptions of chunks of music that we have just heard and that is still in working memory.

Music-listening is a Creative Process

> "To listen is an effort, and just to hear is no merit. A duck hears also." — Igor Stravinsky

Implicit in the discussion in the previous section is the idea that music-listening involves *creating*. Specifically, perceptual structures representing what we have heard are formed in working memory. These are mental objects, not physical and tangible structures, but they are no less real. In this section I elaborate on this idea, starting with some commentary on a relevant-sounding but quite differently motivated project.

Some Thoughts on David Cope's EMI

Any mention of creativity, music, and computers is likely to bring to mind the work of David Cope, so a few words at the outset will be useful to provide some context and to contrast his approach with the one motivating the design of Musicat. Cope's computer program EMI (*Experiments in Musical Intelligence*) was an ambitious and thought-provoking attempt to make a computer compose music in the style of human composers (Cope, 1996). EMI generated pieces of music that sound in many ways like Bach's inventions, Chopin's mazurkas, or Mahler's symphonies — enough so that many musically-educated listeners have had trouble distinguishing human-composed pieces from EMI's imitations. What, then, is left for the science of creativity, if computers can already churn out symphonies in droves? The fact is, though, that incredible challenges remain — and one of Mozart's hobbies, the *Musikalisches Würfelspiel* (musical dice game), provides a clue to what is missing from EMI.

Popular in the 1700's, the musical dice game was a simple algorithmic way of generating musical pieces: dice were used to select pre-composed segments of music at random, which were then strung together in sequence. Many composers of the time (Mozart, for instance) enjoyed composing short segments of music to be used in the process, carefully ensuring that the end of each segment would flow nicely into the beginning of any potential successive segment. Once the short segments had been written, the dice were rolled, fragments were randomly selected, one by one, and the final piece was written down. It is obvious that any creativity one might attach to the result derives from the composition process that occurred *before* the dice-rolling started.

EMI may be seen as using existing human-composed music as input to its own musical dice game — albeit one with quite elaborate rules for combining segments. Roughly speaking, EMI functions as a sophisticated machine for chopping up preexisting music into small fragments and then coherently stitching together those fragments into a larger structure (Cope & Hofstadter, 2001). Indeed, Cope himself repeatedly likens his program to the *Musikalisches Würfelspiel*. Listeners who are very familiar with the source compositions by human composers can often hear original fragments stitched together in EMI's output. (In certain of EMI's pieces, it is far subtler to hear the influence of the source materials, and even highly trained musicians have been fooled into thinking that EMI's output was human-composed.) The computer, however, has no awareness of its own musical output and doesn't even notice if it plagiarizes an entire phrase of a Bach invention — it has no concept of creativity, no mechanism to distinguish exciting new ideas from mere quotations, no inner drive towards genuine novelty, no inner emotional fire. Although it generates music, EMI does so in a manner that seems to me to be radically different from how humans compose. Admittedly, Cope himself might disagree: see (Cope, 1996) for his point of view.

The research program motivating the design of Musicat is committed to the notion of coming to an understanding of human creativity not by frontally attacking the gigantic high-level problem of how a gifted composer dreams up a nocturne or a symphony, but instead via a more humble type of modeling of the fundamental mechanisms of genuine human perception and thought.

Some Thoughts on the Scope of This Project

This research aims to probe mechanisms of creative thought that are fundamental both to music cognition and to cognition in general. Instead of attempting to model the enormously complex process of listening to large works of music or to deal with issues of cross-cultural music perception, this project restricts the modeling task to a particular subclass of music (melodies with simple harmony in the Western tradition). The model proposed here approximates the cognitive processes that take place when a person listens to short, simple-sounding fragments of music. This restriction can be surprisingly helpful — not only does it bring the modeling task into the realm of the tractable, but it also forces the model to focus on core perceptual issues.

Some Thoughts on Creative Music Perception

Perception is an active process in which the brain interprets sensory stimuli, such as sound waves, and transforms the raw input into new mental structures that are more suitable — indeed, essential — for the activities of high-level cognition. Far from being a passive, mechanical algorithm hard-wired into our brains from birth, this process of interpretation is dynamic, involving a combination of bottom-up and top-down processing. By necessity, the interpretive process ignores a huge number of details and focuses attention on the most important aspects of a situation — otherwise, we would we overwhelmed with sensory information (just think of the huge amount of detail available to our senses in, say, a crowded restaurant: the number of objects we can see, things we can smell, or sounds we can hear is nearly uncountable). Such perception is subjective and creative. A central tenet of the philosophy underlying this work is that that exactly the same processes that give rise to the high-level *generation* of creative thoughts, ideas, and works of art also drive the creative *perception* of sensory input. The notion of creativity residing in the humble act of perception might seem, at first glance, far-fetched, but the view advocated here is that the essence of creativity is evident both in "mere" perception and in activities that we standardly think of as creative.

Creativity is an intimate part of music perception in particular because listening to music requires a listener to form subjective internal mental representations of the music. Two different people listening to the exact same recording of a piece can have radically different experiences, as if they each heard completely different music — consider a European hearing Balinese music for the first time, and a Balinese person hearing the same piece. The raw sensory input is the same for both listeners, but their subjective experiences and their internal representations of the heard music would be quite different. We need not go as far as Bali to see how listeners differ, however. Even two American listeners will likely hear the same piece in quite different ways: a person who loves romantic-era music will certainly understand Chopin in a different way from a person who loves heavy metal. Furthermore, a classical musician familiar with baroque music might hear Chopin differently from the way a musician familiar with romantic music would. In the extreme case, even two different professional concert pianists may hear the same Chopin piece differently — just think of how different two different performers' interpretations of the same piece can be. Although performing and listening are distinct talents, I suspect that some of the variability in interpretation stems from the variability performers' own idiosyncratic ways of listening. Generating representations is a creative act, as is demonstrated by the uniqueness of each listener's experience. These internal representations also induce expectations about what the music will do next. This act of generating expectations is closely related to the process of music composition, and is obviously creative.

Music-listening is Not a Universal Talent

> "When I speak of the gifted listener, I am thinking of the non-musician primarily, of the listener who intends to retain his amateur status. It is the thought of just such a listener that excites the composer in me." — Aaron Copland

I have argued above that music-listening is a dynamic and creative process. In this section I go further, claiming that people have differing degrees of talent for listening. Musicat's listening is extremely primitive compared to almost all human listening, but one should bear in mind that Musicat's ultimate goal is to model the listening performance of what Copland called a "gifted listener" in the quote above. Otherwise, we could imagine that much simpler programs could possibly model "listening".

We are surrounded by music in contemporary society. I am listening to music on my headphones as I write these words. We routinely listen to music streamed in real time from the Internet. We can carry countless of gigabytes of music on our iPods to listen to as we go jogging or driving. We can't avoid the nearly constant stream of background music playing in stores or restaurants. Not long ago, I was in a restaurant where the music from a live band in one room was competing with prerecorded music that was playing in the next room. What a cacophony! And yet such things are quite commonplace today.

Inundated with music as we are, we might be inclined to think of everyone as an expert listener — with such intense exposure to music all the time, it would seem plausible that we'd all be really good at listening. However, the kind of listening that I am interested in is not automatic, nor is it even possible for everyone. As an analogy, consider photography. In the early days, people thought that photography involved simply clicking a button to take a picture — the camera would do all the work. In contrast, in the case of arts such as drawing or painting, the artist was clearly putting in a lot of effort. But as we understand well today, photography too is an art, requiring thoughtfulness, talent, and work, although the work of a photographer might be less evident as such than that of a painter. A photographer is also *creating* by composing the scene, selecting the particular position of the camera, arranging for the right lighting, making choices of lens and aperture, setting the focus on a particular part of the scene, and so on and so forth. The photographer, above all, is thoughtfully considering the visual scene and making choices about how to transmit a particular view of the scene to another person through an image. Just as the photographer does more than simply click a button to take a photo, the listener must actively consider music, decide which aspects of the music to focus on, and create internal representations of the music that is heard — it is much more than simply clicking a button or putting on a pair of headphones.

Listeners perform little externally-apparent "work", but people differ greatly in their skill at listening. Some years ago, it was not uncommon to hear a person modestly say, "I have a tin ear" or "I'm tone-deaf" to indicate a lack of music-listening skill. Although these phrases have somewhat fallen out of fashion, it stands to reason that people today still vary in their levels of listening skill.

Listening skill varies among people, because listening is quite complex. I want to be clear up-front that Musicat is not a great listener when compared with most people. The program, as will be seen, still fails to understand some things that almost all human listeners (at least those who grew up listening to Western tonal music) would intuitively understand. At the same time, though, it "hears", with its "silicon ears", a few rather sophisticated things that might go unnoticed by the tone-deaf "tin ears" among us.

Overview of this Dissertation

This dissertation could be seen as consisting of four parts, as follows:

1. Four initial chapters (including this one) set Musicat in the context of other work, explain the domain of music-listening that Musicat is concerned with, and discuss why analogy-making is central to this work.
2. The next three chapters give numerous detailed examples of Musicat in action, showing its listening performances on a variety of melodies.
3. The next two chapters describe the design of the Musicat program, both in its present form and in earlier incarnations.
4. The final chapter includes a discussion of Musicat's current state as a listener, describing what it does well and what it does badly, and suggesting directions for future work.

There are also three appendices providing bibliographic references, a description of a pilot study, and some preliminary quantitative results comparing Musicat to another model.

Whither Musicat?

In his Norton Lectures in 1973, Leonard Bernstein asked the question "Whither music?" and more specifically "Whither music in our time?" For Bernstein, this was a crucial question — it was his version of Charles Ives' musical "Unanswered Question".

I will give some tentative and speculative answers to my own question "Whither Musicat?" at the end of this thesis, in Chapter 10. But for now, harkening back to Ives, I'll leave the question unanswered. It is hard to say where Musicat is headed, because Musicat still has a long way to go before it will come anywhere close to the complexities of human listening; today's version is just a tiny first step towards understanding listening as done by people. But it's a step in the right direction, in my extremely biased opinion.