

Implications from Music Generation for Music Appreciation

Amy K. Hoover, Paul A. Szerlip, and Kenneth O. Stanley

Department of Electrical Engineering and Computer Science

University of Central Florida

Orlando, FL 32816-2362 USA

{ahoover@eecs.ucf.edu,paul.szerlip@gmail.com,kstanley@eecs.ucf.edu}

Abstract

This position paper argues that fundamental principles that are exploited to achieve effective music generation can also shed light on the elusive question of why humans appreciate music, and which music is easiest to appreciate. In particular, we highlight the key principle behind an existing approach to assisted accompaniment generation called *functional scaffolding for musical composition* (FSMC). In this approach, accompaniment is generated as a *function* of the preexisting parts. The success of this idea at generating plausible accompaniment according to studies with human participants suggests that perceiving a *functional relationship* among parts in a composition may be essential to the appreciation of music in general. This insight is intriguing because it can help to explain without any appeal to traditional music theory why humans with no knowledge or training in music can nevertheless find satisfaction in coherent musical structure.

Introduction

Among the most fundamental questions on the human experience of music is why we appreciate it so universally and what makes some pieces more appealing than others (Hanslick, 1891; Sacks, 2008; Frith, 2004; Gracyk, 1996). There are many possible approaches to addressing these questions, from studies of expectation fulfillment (Huron, 2006; Schmuckler, 1989; Pearce and Wiggins, 2012; Abdallah and Plumbley, 2009) to cultural factors (Balkwill and Thompson, 1999; Peddie, 2006). Our aim in this paper is to propose an alternative route to addressing the fundamental basis for music appreciation, by beginning with an approach to *music generation* and from its mechanics drawing implications for at least one key underlying ingredient in the appreciation of music. The motivation is that the process of designing an effective music generator implicitly forces the designer to confront the basis of music appreciation as well. After all, a music generator is little use if its products are not appealing.

Particularly revealing would be a simple principle that almost always can be applied. The simpler such a principle, the more plausible that it might explain some aspect of music appreciation. One approach to assisted music generation based on such a simple principle is called *functional scaffolding for musical composition* (FSMC) (Hoover and Stanley, 2009; Hoover, Szerlip, and Stanley, 2011b,a; Hoover et

al., 2012). Our position is that the principle at the heart of this approach, initially conceived as a basis for generating accompaniment, offers a unique hint at the machinery behind human musical appreciation. In this way, it can contribute to explaining in part both when and why humans appreciate music.

Functional Scaffolding for Musical Composition (FSMC)

The FSMC approach is based on the insight that music is at heart a pattern of notes played over time with some regularity. As a result, one way to conceptualize music is as a *function of time*. Formally, for any musical voice, the pattern of pitches and the pattern of durations and rests can be expressed together as a vector function of time $f(t)$ that outputs both pitch and rhythm information. In practice, to generate a sequence of notes, f could be queried at every time t and the complete output sequence would constitute the pattern. The parts played by each instrument in an ensemble piece could also be output simultaneously by such a function.

This perspective is helpful for music generation when combined with the insight that all the instrumental sequences (i.e. each track) in a single piece must be somehow related to each other. For example, in a popular rock piece, the drum pattern, say $d(t)$, typically establishes the rhythm for the rest of the piece. Therefore, the bass pattern, $b(t)$, which helps structure the harmonic form, will by necessity depend in some way upon the drum pattern. This idea of relatedness between parts can be expressed more formally by saying that the bass pattern is a *function* of the drum pattern, which can be expressed by a function h that relates $b(t)$ to $d(t)$: $b(t) = h(d(t))$. Building on the drum and bass patterns, vocalists and other instrumental parts can then explore more complicated melodic patterns that are themselves also related to the established rhythmic and chord patterns. It follows then that not only can each of these instrumental parts be represented as a function of time, but that they are indeed each functions of *each other*.

Beyond just observations, these insights imply a practical opportunity for generating musical accompaniment. By casting instrumental parts as functions of each other, the problem of accompaniment is illuminated in a new light:



Figure 1: **Representing and Searching for Accompaniments with FSMC.** The function $f(t)$, which is depicted by a piano keyboard, represents the human composition called the *scaffold*, from which the computer-generated accompaniments are created. A possible such accompaniment, $g(t)$, is shown atop and depicted by the image of a computer. Each accompaniment is internally represented by a helper function, $h(f(t))$, which is represented by a special type of artificial neural network (ANN) called a compositional pattern producing network (CPPN). Like ANNs, CPPNs can theoretically approximate any continuous function. Thus these CPPNs represent h , which transforms the scaffold into an accompaniment.

Given an existing part $f(t)$, the problem of formulating an appealing accompaniment becomes the problem of searching for accompaniment $g(t)$ such that $g(t)$ complements $f(t)$. Yet while applying a search algorithm directly to finding such a function $g(t)$ would be difficult because the search space is vast, instead the search can be significantly constrained by searching for $h(f(t))$, as depicted in figure 1. The major benefit of this approach is that because h is a function of the part it will accompany, it cannot help but follow to some extent its contours. Therefore, the idea for generating accompaniment in FSMC is to search with the help of a human user for a function $h(f(t))$, where $f(t)$ is a preexisting part or *scaffold*. By searching for a transforming function instead of an explicit sequence of notes, the plausibility of output accompaniments is enhanced. In effect, $f(t)$ provides the *functional scaffolding* for the accompaniment.

The idea in FSMC that searching for $h(f(t))$ can yield plausible accompaniment to $f(t)$ can be exploited in practice by programming a search algorithm to explore possible variations of the function h . In fact, this approach has been tested extensively in practice through an implementation called MaestroGenesis (<http://maestrogenesis.org/>), whose results have been reported in a number of publications (Hoover, Szerlip, and Stanley, 2011b,a; Hoover et al., 2012). In MaestroGenesis, the function h is represented by a special kind of artificial neural network called a compositional pattern producing network (CPPN). A population of candidate CPPNs is *evolved* interactively by allowing a human user to direct the search algorithm by picking his or her favorite candidate accompaniments to produce the offspring for the next generation. Thus the *representation* of the transforming function is a CPPN (which is a kind of function approximator) and the *search algorithm* is interactive evolution (which is an evolutionary algorithm guided by a human; Takagi, 2001). A full

technical description is given in Hoover, Szerlip, and Stanley (2011a), Hoover, Szerlip, and Stanley (2011b), and Hoover et al. (2012).

Interestingly, listener study results from FSMC-generated music showed that musical pieces with accompaniments that were generated purely through functional relationships were indistinguishable from fully human composed pieces (Hoover, Szerlip, and Stanley, 2011a). In fact, some fully human composed pieces were rated more mechanical-sounding than those that were only partially human composed. Similarly positive results were also reported in other studies (Hoover and Stanley, 2009; Hoover, Szerlip, and Stanley, 2011b,a; Hoover et al., 2012). Although more variation in the initial human composition (i.e. a polyphonic versus monophonic scaffold) provides more richness from which to work, as Hoover et al. (2012) show, plausible accompaniments can nevertheless be generated from as little as a single monophonic starting melody. Furthermore, often in MaestroGenesis even the first generation of candidate accompaniments, which are randomly-generated CPPNs, sounds plausible because the functional relationship ensures at least some relationship between the scaffold and its accompaniment (Hoover and Stanley, 2009).

From Music Generation to Music Appreciation

These results are of course relevant to progress in music generation, but they hint at a deeper implication. In particular, it is notable that MaestroGenesis (and FSMC behind it) has *almost no musical knowledge* programmed into it. In fact, the only real musical rule in the program is that CPPN outputs are forced to be interpreted as notes within the key of the scaffold track. Aside from that, MaestroGenesis has no knowledge of chords, rhythm, progression, melody, har-

mony, dissonance, style, genre, or anything else that a typical music generator might have (Simon, Morris, and Basu, 2008; Chuan, 2009; Ebcioğlu, 1990). It thus relies almost entirely on the functional relationship between the scaffold and the accompaniment to achieve plausibility. In effect, the functional relationship causes the accompaniment to *inherit* the gross structure of the scaffold, thereby endowing it with many of the same aesthetic properties. Thus the key observation behind this position paper is that establishing such a functional relationship between different parts of a song seems to be *sufficient on its own* to achieve plausible musical structure.

This observation is intriguing because it implies a hypothesis about the nature of musical appreciation: If a functional relationship alone is sufficient to achieve musical plausibility in the experience of human listeners, then perhaps musical appreciation itself is at least in part the result of perceiving a functional relationship between different parts of a composition. That is, functional relationships, which are *mathematical properties* of patterns that do not require any specific musical knowledge to perceive, could explain why listeners without any musical training or expertise nevertheless experience and appreciate music and separate it firmly from cacophony. In effect the human is appreciating the functional relationship that binds different parts of a composition together.

If true, this hypothesis can explain to some extent when humans will or will not appreciate a composition. For example, the harder it is to perceive how one part is functionally related to another, the less pleasing that piece may be. Such functional relationships are potentially perceived not only between different instrumental parts or tracks, but also within a single instrumental part played over time. That is, if a functional relationship can be perceived between an earlier sequence of notes and a later one, then the entire sequence may succeed as musically plausible or even appealing.

At the same time, it may also explain why some compositions are more difficult to enjoy. For example, some research in computer music explores the sonification of non-auditory data (Cope, 2005; Park et al., 2010; Vickers, 2005). Typically, the user inputs semi-random data to a computer model (e.g. cellular automata, swarms, etc.) that outputs music. While the output is a function of the input, because the initial seed does not stem from inherently musical events, the outputs are often difficult for non-creators to immediately understand. However, as these systems develop, composers begin to build musical frames for anticipation and expectation. While there can certainly be beauty in such pieces, the audience needs some familiarity, like the composer, with the style to begin to perceive the important relationships.

Though perhaps not explicitly, composers have long intuited the importance of incorporating functional relationships into compositions. Not only are musical lines regularly translated, inverted, and reflected, but some logarithmic and modular transformations (and set theory concepts) pre-date the mathematical formalisms themselves (Risset, 2002; Harkleroad, 2006). The implicit nature of the composers' insight is that people appreciate these functional transformations within a "relatable" musical context.

In fact, much of compositional musical theory was developed to produce consistent aesthetic results (Payne, 1995; Christensen, 2002). By following certain heuristics and established patterns, composers ensure that pieces fit within particular styles and genres. Such heuristics generally encompass narrow sets of phenomena, e.g. waltzes, counterpoint, jigs, etc. The hypothesis that functional relationships provide a general principle for musical appreciation provides a unifying perspective for all such disparate stylistic conventions: At some level, all of them ultimately establish some kind of functional relationship among the parts of a composition.

Furthermore, this perspective suggests that as long as they are perceptible (i.e. not so complex as to sound cacophonous), relatively simple functions likely exist that generate relationships among parts that are aesthetically appealing yet not related to any genre, rule, or heuristic currently taught or even yet conceived. For example, music generated with FSMC exhibits a range of complexity, suggesting little restriction on the type of function necessary to create plausible accompaniments. Some of the most appealing accompaniments are generated from very simple relationships (Hoover, Szerlip, and Stanley, 2011b,a), while sometimes more complex relationships between melody and percussion are also appreciated by listeners (Hoover and Stanley, 2009). To some extent this theory thereby suggests without any other musical theory when breaking the rules might be appealing and when it might not; as long as a functional relationship among parts can still be perceived, the average listener will not necessarily react negatively to breaking established conventions.

Sometimes it can also take a while to habituate to styles or genres that do not follow conventional rules. For example, atonal pieces can be difficult to enjoy for the uninitiated. Interestingly, the functional hypothesis provides a potential insight into why such unconventional styles can become appealing with experience. The explanation is that initially the functional relationships among different parts in such an unconventional context are difficult to perceive because the relationships are both complex and unfamiliar. Therefore, the brain initially struggles to identify the functional relationship. However, over time, repeated exposure familiarizes the listener with the kinds of transformations that are typical in the new context, such that eventually the brain can pick out functional relationships that once were too complex to perceive. At that point, the music becomes possible again to appreciate. In fact, as noted by Huron (2006), the patterns associated with a particular style are designed to elicit emotions by playing on the listeners' expectations. Those expectations can be viewed as mediated by the kinds of functional relationships with which the listener is familiar.

Music theory provides many heuristics for composing plausible types of music like fugues or walking bass lines. But as any musician knows, simply following such rules without the elusive element of inspiration results in plausible yet dry-sounding pieces. A good musician must know the standard rules for composition and also when to break them, but the problem of when to break the rules in music theory is less understood than the standard rules for composition. The

insight that a rule is well-broken if it still preserves a perceptible functional relationship provides a possible direction for studying this issue further.

Conclusions

This type of general theory follows directly from taking a minimalist approach to music generation. Approaches that rely on acquiring or enumerating all the complexities of music of certain types or composers of certain types, such as through statistical inference (Rhodes, Lewis, and Müllensiefen, 2009; Kitani and Koike, 2010) or grammatical rules (Holtzman, 1981; McCormack, 1996), cannot probe the possibility of deeper underlying principles than the rules that are apparent at the surface. In contrast, FSMC and MaestroGenesis took the minimalist approach to music generation by predicating everything only on functional relationships. While a potential criticism of such an approach is that it is too simplistic to capture all the subtlety of sophisticated musical composition, its benefit in a scientific context is that it *isolates* a single phenomenon so that the full implication of that phenomenon can be tested. The result is a simple hypothesis that reduces musical theory to a mathematical principle, i.e. perceiving functional relationships, that can plausibly be appreciated even by listeners without musical training. It also becomes a tool for music generation, as in MaestroGenesis, that does not require enumerating complex rules. While functional relationships need not constitute the entire explanation for all musical appreciation, they are an appealing ingredient because of their simplicity and possibility for future study – they suggest that within the mind of a composer perhaps at some level such a function is realized as the overall pattern of a musical piece is first conceived.

In a broader context, explaining the appreciation of music through perceiving functional relationships also connects musical appreciation to non-musical aesthetics. After all, across the spectrum of art, architecture, and even human beauty, symmetry, repetition, and variation on a theme are paramount. It is notable that all such regularities ultimately reduce to one instance of a pattern being functionally related to another. Given that we appreciate such relationships in so many spheres of our experience, that music too would draw from such an affinity follows elegantly.

Acknowledgments

This work was supported in part by the National Science Foundation under grant no. IIS-1002507 and also by a NSF Graduate Research Fellowship. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors(s) and do not necessarily reflect the views of the National Science Foundation.

References

- Abdallah, S., and Plumbley, M. 2009. Information dynamics: Patterns of expectation and surprise in the perception of music. *Connection Science* 21(2).
- Balkwill, L.-L., and Thompson, W. F. 1999. A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music Perception* 43–64.
- Christensen, T. 2002. *The Cambridge History of Western Music Theory*. Cambridge University Press.
- Chuan, C.-H. 2009. Supporting compositional creativity using automatic style-specific accompaniment. In *Proc. of the CHI Computational Creativity Support Workshop*.
- Cope, D. 2005. *Computer Models of Musical Creativity*.
- Ebcioğlu, K. 1990. An expert system for harmonizing chorales in the style of J.S. Bach. *Journal of Logic Programming* 145–185.
- Frith, S. 2004. What is bad music? In Washburne, C., and Derno, M., eds., *Bad Music: The Music We Love to Hate*. Routledge. 15–36.
- Gracyk, T. 1996. *Rhythm and Noise: An Aesthetics of Rock*. IB Tauris.
- Hanslick, E. 1891. *On the Musically Beautiful: A Contribution Towards the Revision of the Aesthetics of Music*. Novello Ewer and Company.
- Harkleroad, L. 2006. *The Math Behind the Music*. Outlooks. Cambridge University Press.
- Holtzman, S. R. 1981. Using generative grammars for music composition. *Computer Music Journal* 5(1):51–64.
- Hoover, A. K., and Stanley, K. O. 2009. Exploiting functional relationships in musical composition. *Connection Science Special Issue on Music, Brain, & Cognition* 21(2):227–251.
- Hoover, A. K.; Szerlip, P. A.; Norton, M. E.; Brindle, T. A.; Merritt, Z.; and Stanley, K. O. 2012. Generating a complete multipart musical composition from a single monophonic melody with functional scaffolding. In *To appear in: Proceedings of the Third International Conference on Computational Creativity (ICCC-2012, Dublin, Ireland)*.
- Hoover, A. K.; Szerlip, P. A.; and Stanley, K. O. 2011a. Generating musical accompaniment through functional scaffolding. In *Proceedings of the Eighth Sound and Music Computing Conference (SMC 2011)*.
- Hoover, A. K.; Szerlip, P. A.; and Stanley, K. O. 2011b. Interactively evolving harmonies through functional scaffolding. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2011)*. New York, NY: The Association for Computing Machinery.
- Huron, D. 2006. *Sweet Anticipation: Music and the Psychology of Expectation*.
- Kitani, K. M., and Koike, H. 2010. Improv-generator: Online grammatical induction for on-the-fly improvisation accompaniment. In *Proceedings of the 2010 Conference on New Interfaces for Musical Expression (NIME 2010)*.
- McCormack, J. 1996. Grammar based music composition. *Complex Systems* 96:321–336.
- Park, S.; Kim, S.; Lee, S.; and Yeo, W. S. 2010. Composition with path: Musical sonification of geo-referenced data with online map interface. In *Proceedings of the International Computer Music Conference*.

- Payne, S. K. D. 1995. *Tonal Harmony: With an Introduction to Twentieth*. McGraw-Hill.
- Pearce, M. T., and Wiggins, G. A. 2012. Auditory expectation: The information dynamics of music perception and cognition. *Topics in Cognitive Science* 4(4):625–652.
- Peddie, I. 2006. *The Resisting Muse: Popular Music and Social Protest*. Ashgate Publishing Company.
- Rhodes, C.; Lewis, D.; and Müllensiefen, D. 2009. Bayesian model selection for harmonic labelling. *Mathematics and Computation in Music* 107–116.
- Risset, J.-C. 2002. Computing musical sound. In Assayag, G.; Feichtinger, H. G.; and Rodrigues, J. F., eds., *Mathematic and Music: A Diderot Mathematical Forum*. Springer-Verlag. chapter 13.
- Sacks, O. 2008. *Musicophilia: Tales of Music and the Brain*. Vintage Canada.
- Schmuckler, M. A. 1989. Expectation in music: Investigation of melodic and harmonic processes. *Music Perception: An Interdisciplinary Journal* 7(2):109–149.
- Simon, I.; Morris, D.; and Basu, S. 2008. Mysong: Automatic accompaniment generation for vocal melodies. In *Proc. of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems*, 725–734. ACM.
- Takagi, H. 2001. Interactive evolutionary computation: Fusion of the capacities of EC optimization and human evaluation. *Proceedings of the IEEE* 89(9):1275–1296.
- Vickers, P. 2005. Ars informatica–ars electronica: Improving sonification aesthetics. In *Understanding and Designing for Aesthetic Experience Workshop at HCI 2005 The 19th British HCI Group Annual Conference*.