

Generating a Complete Multipart Musical Composition from a Single Monophonic Melody with Functional Scaffolding

Amy K. Hoover, Paul A. Szerlip, Marie E. Norton, Trevor A. Brindle,
Zachary Merritt, 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, marie.norton@knights.ucf.edu, tabrindle@gmail.com,
zmmerritt@gmail.com, kstanley@eecs.ucf.edu }

Abstract

This paper advances the state of the art for a computer-assisted approach to music generation called *functional scaffolding for musical composition* (FSMC), whose representation facilitates creative combination, exploration, and transformation of musical concepts. Music in FSMC is represented as a functional relationship between an existing human composition, or *scaffold*, and a generated accompaniment. This relationship is encoded by a type of artificial neural network called a compositional pattern producing network (CPPN). A human user without any musical expertise can then explore how accompaniment should relate to the scaffold through an interactive evolutionary process akin to animal breeding. While the power of such a functional representation has previously been shown to constrain the search to plausible accompaniments, this study goes further by showing that the user can tailor complete multipart arrangements from only a single original monophonic track provided by the user, thus enabling creativity without the need for musical expertise.

Introduction

Among the most important functions of any approach to enhancing human creativity is what Boden (2004) terms *transformational creativity*. That is, key creative obstacles faced by human artists and musicians are the implicit constraints acquired over a lifetime that shape search space structure. By offering an instance of the search space (e.g. of musical accompaniments) with a radically different structure, a creativity-enhancing program can potentially liberate the human to discover unrealized possibilities. In effect, the familiar space of the human artist is *transformed* into a new structure intrinsic to the program. Once the user is exposed to this new world, as a practical matter the program must provide to the user the ability to *explore* and *combine* concepts within the newly-conceived search space, which corresponds to Boden's *combinatorial* and *exploratory* classes of creativity (Boden, 2004). That way, the user experiences a rich and complete creative process within a space that was heretofore inconceivable.

The danger with transformational creativity in computational settings is that breaking hard-learned rules may feel unnatural and thereby unsatisfying (Boden, 2007). Any attempt to facilitate transformational creativity should respect the relationships between key artistic elements even as they are presented in a new light. Thus for a given domain, such as musical accompaniment, a delicate balance must be struck between unfettered novelty and respect for essential structure.

Many approaches to generating music focus on producing a natural sound at the cost of restricting creative exploration. Because structure is emphasized, the musical space is defined by rules that constrain the results to different styles and genres (Todd and Werner, 1999; Chuan, 2009; Cope, 1987). The necessity for a priori rules potentially facilitates the combination of musical structures or exploration of the defined space, but precludes transformational outcomes.

In contrast, musical structures in the approach examined in this paper, *functional scaffolding for musical composition* (FSMC), are defined as the very functions that relate one part of a piece to another, thereby enabling satisfying transformational creativity (Hoover, Szerlip, and Stanley, 2011a,b). Based on the idea that music can be represented as a function of time, FSMC inputs a simple, isolated musical idea into a function that outputs accompaniment that respects the structure of the original piece. The function is represented as a special type of artificial neural network called a compositional pattern producing network (CPPN). In practice, the CPPN inputs existing music and outputs accompaniment. The user-guided creative exploration itself is facilitated by an interactive evolutionary technique that in effect allows the user to *breed* the key functional relationships that yield accompaniment, which supports both combinatorial and exploratory creativity (Boden, 2004) through the crossover and mutation operators present in evolutionary algorithms. By representing music as *relationships* between parts of a multipart composition, FSMC creates a new formalism for a musical space that transforms its structure for the user while still respecting its fundamental constraints.

Hoover, Szerlip, and Stanley (2011a,b) showed that FSMC can produce accompaniments that are indis-

tinguishable by listeners from fully human-composed pieces. However, the accompaniment in these studies was only a single monophonic instrument, leaving open the key question of whether a user with little or no musical expertise can perhaps generate an entire multipart arrangement with this technology from just a single-instrument monophonic starting melody. If that were possible, then anyone with only the ability to conceive a single, monophonic melody could in principle expand it into a complete multilayered musical product, thereby enhancing the creative potential of millions of amateur musicians who possess inspiration but not the expertise to realize it. This paper demonstrates that FSMC indeed makes such achievement possible.

Background

This section relates FSMC to traditional approaches to automated composition and previous creativity-enhancing techniques.

Automatic Music Generation

Many musical representations have been proposed before FSMC, although their focus is not necessarily on representing the functional relationship between parts. For example, from long before FSMC, Holtzman (1980) creates a musical grammar that generates harp solos based on the physical limitations imposed on harp performers. Similarly, Cope (1987) derives grammars from the linguistic principles of haiku to generate music in a particular style. These examples and other grammar-based systems are predicated on the idea that music follows grammatical rules and thus by modeling musical relationships as grammars, they are representing the important structures of music (Roads, 1979; McCormack, 1996). While grammars can produce a natural sound, deciding which aspects of musical structure should be represented by them is often difficult and ad hoc (Kippen and Bel, 1992; Marsden, 2000).

Impro-Visor helps users create monophonic jazz solos by automatically composing any number of measures in the style of famous jazz artists (Keller et al., 2006). Styles are represented as grammars that the user can invoke to complete compositions. Creativity-enhancement in Impro-Visor occurs through the interaction of the user's own writing and the program's suggestions. When users have difficulty elaborating musical concepts, they can access predictions of how famous musicians would approach the problem within the context of the current composition. By first learning different professional compositional techniques, students can then begin developing their own personal styles. While Impro-Visor is an innovative tool for teaching jazz styles to experienced musicians, it focuses on emulating prior musicians over exploration.

Enhancing Creativity in Music Composition

A problem with traditional approaches to music composition is that standard representations can poten-

tially limit creative exploration. For instance, MySong generates chord-based accompaniment for a vocal piece from hidden Markov models (Simon, Morris, and Basu, 2008). Users select any vocal piece and MySong outputs accompaniment based on a transition table, a weighting factor that permits greater deviation from the table, and musical style (e.g. rock, big band). MySong thus allows users to create accompaniment for their own melodies in a variety of different predefined styles from which users cannot deviate. Zicarelli (1987) describes an early interactive composition program, Jam Factory, that improvises on human-provided MIDI inputs from rules represented in transition tables. Users manipulate the output in several ways including the probability distributions of eight different transition tables; there are four each for both rhythm and pitch. Users are provided more creative control in designing and consulting the transition tables, but the increased flexibility results in unnatural outputs that thereby limit the utility of the main algorithms (Zicarelli, 2002). The approach described by Chuan (2009) balances user control by training transition tables based on only a few user-provided examples. The tables then reflect the "style" inherent in the examples and can generate chord-based accompaniment for a user's own piece. While each of these systems offers users varied levels of control, rule manipulation alone may not be sufficient to access all three forms of creativity described by Boden (2004). For example, the representations cannot easily combine musical ideas or transform the musical space (due to inherent rule restrictions).

Alternatively, most interactive evolutionary computation (IEC) (Takagi, 2001) approaches facilitate creativity through the evolutionary operators of crossover and mutation, and require human involvement in the creative process. In GenJam a human player and computer "trade fours," a process whereby the human plays four measures and the computer "answers" them with four measures of its own (Biles, 1998). Musical propositions are mutated and combined into candidates that the user rates as good or bad. Similarly, Jacob (1995) introduces a system in which human users rate, combine, and explore musical candidates at three different levels of the composition process and Ralley (1995) generates melodies by creating a population from mutations of a provided user input. Finally, CACIE creates atonal pieces by concatenating musical phrases as they are generated over time (Ando and Iba, 2007). Each phrase is represented as a tree structure that users can interactively evolve or directly manipulate. However, most such systems impose explicit musical rules conceived by the developer to constrain the search spaces of possible accompaniment, thus narrowing the potential for discovery.

Previous Work in FSMC

The FSMC approach in this paper is based on previous work by Hoover, Szerlip, and Stanley (2011a,b), who focused on evolving a single monophonic accompa-

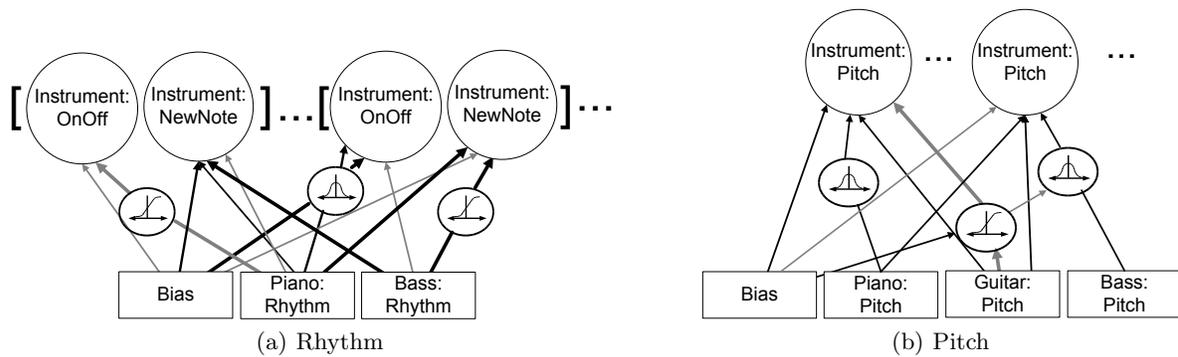


Figure 1: **How CPPNs Compute a Function of the Input Scaffold.** The rhythm CPPN in (a) and pitch CPPN in (b) together form the accompaniments of FSMC. The inputs to the CPPNs are the scaffold rhythms and pitches for the respective networks and the outputs indicate the accompaniment rhythms and pitches. Each rhythm network has two outputs: OnOff and NewNote. The OnOff node controls volume and whether or not a note is played. The NewNote node indicates whether a note is re-voiced or sustained at the current tick. If OnOff indicates a rest, the NewNote node is ignored. The pitch CPPN output decides what pitch the accompaniment should play at that particular tick. The internal topologies of these networks, which encode the functions they perform, change over evolution. The functions within each node depict that a CPPN can include more than one activation function, such as Gaussian and sigmoid functions. Two monophonic accompaniment outputs are depicted, but the number of instruments a CPPN can output is unlimited. The number of input instruments also can vary.

niment for a multipart MIDI. These accompaniments are generated through two functions, one each for pitch and rhythm, that are represented as compositional pattern producing network (CPPNs), a special type of artificial neural network (ANN). CPPNs can evolve to assume an arbitrary topology wherein each neuron is assigned one of several activation functions. Through IEC, users explore the range of accompaniments with NeuroEvolution of Augmenting Topologies (NEAT), a method for growing and mutating CPPNs (Stanley and Miikkulainen, 2002). Unlike traditional ANN learning, NEAT is a policy search method, i.e. it explores accompaniment possibilities rather than optimizing toward a target. While existing songs with generated accompaniments were indistinguishable in a listener study from fully-composed human pieces, the real achievement for this approach would be to help the user generate entire polyphonic and multi-instrument accompaniment from just a single voice of melody (Hoover, Szerlip, and Stanley, 2011a). This paper realizes this vision.

Approach: Extending Functional Scaffolding for Music Composition

This section extends the original FSMC approach, which only evolved a single monophonic accompaniment (Hoover, Szerlip, and Stanley, 2011a,b). It explains the core principles of the approach and how they are applied to producing multipart accompaniments.

Defining the Musical Space

A crucial aspect of any creativity-enhancing approach for music composition is first to define the musical space. Users can help define this space in FSMC by first

selecting a musical starting point, i.e. the monophonic melody or *scaffold*. Initial scaffolds can be composed in any style and if they are only single monophonic parts as in this paper, they can be composed by users within a wide range of musical skill and expertise. The main insight behind the representation in FSMC is that a robust space of accompaniments can be created with only this initial scaffold. Because of the relationship of different accompaniment parts to the scaffold and therefore to each other, the space is easily created and explored.

Each instrument part in the accompaniment is the result of two separate functions that independently relate rhythmic and pitch information in the scaffold (i.e. the inputs) to the generated accompaniment. Depicted in figure 1, these functions are represented as CPPNs, the special type of ANN described in the background (Stanley, 2007). As figure 1 shows, multiple inputs can be sent to the output and many different instruments can be represented by the same CPPN. CPPNs incrementally grow through the NEAT method, which means they can in principle evolve to represent *any* function (Stanley and Miikkulainen, 2002; Cybenko, 1989). Together, the rhythmic and pitch CPPNs that will be evolved through NEAT define the musical space that the user can manipulate. In effect, pitch information from the scaffold is fed into the pitch CPPN at the same time as rhythmic information is fed into the rhythm CPPN. Both CPPNs then output how the accompaniment should behave in response. That way, they compute a function of the scaffold.

Accompaniments are divided into a series of discrete time intervals called *ticks* that are concatenated together to form an entire piece. Each tick typically represents the length of an eighth note, but this division can

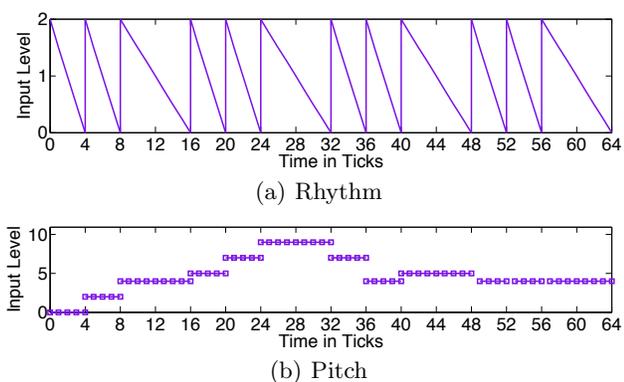


Figure 2: **Input Representation.** The spike-decay representation for rhythmic inputs is shown in (a) and the pitch representation is in (b). Rhythm is encoded as a set of decaying spikes that convey the duration of each note. Because the CPPN sees *where* within each spike it is at any given tick, in effect it can synchronize its outputs with the timing of the input notes. Pitch on the other hand is input simply as the current note at the present tick.

be altered through an interface. Outputs are gathered from both the rhythmic and pitch CPPNs at each tick that are combined to determine the accompaniment at that tick. As shown in figure 1a, the two outputs of the rhythm network for each line of accompaniment are *On/Off*, which indicates whether a note or rest is played and its volume, and *NewNote*, which indicates whether or not to sustain the previous note. The single *pitch* output for each line of accompaniment in figure 1b determines instrument pitch at the current tick relative to a user-specified key.

To produce the outputs, rhythmic and pitch information from the scaffold is sent to the CPPN at each tick. The continuous-time graph in figure 2a illustrates how rhythmic information in the scaffold is sent to the CPPN. When a note strikes, it is represented as a maximum input level that decays linearly over time (i.e. over ticks) until the note ends. At the same tick, pitch information on the current note is input as a pitch class into the pitch CPPN (figure 2b). That is, two C notes in different octaves (e.g. C4 and C5) are not distinguished.

The sound of instruments in FSMC can be altered through instrument choice or key. A user can pick any of 128 pitched MIDI instruments and can request any key. Once a user decides from what preexisting piece the scaffold is provided and the output instruments most appropriate for the piece, candidate CPPNs can be generated, thus establishing the musical space of accompaniments. The theory behind this approach is that by exploring the potential *relationships* between scaffolds and their accompaniments (as opposed to exploring direct representations of the accompaniment itself), the user is constrained to a space in which candidate accompaniments are almost all likely *coherent* with re-

spect to the scaffold. The next section describes how users can combine, explore, and transform this space to harness their own musical creativity.

Navigating the Musical Space



Figure 3: **Program Interface.** This screenshot of the program (called MaestroGenesis) that implements FSMC shows accompaniments for a melody input by the user. The instrument output is currently set to Grand Piano on the left-hand side, but can be changed through a menu. Accompaniments are represented as horizontal bars and are named by their ID. The user selects his or her favorite and then requests a new generation of candidates

Exploration of musical space in FSMC begins with the presentation to the user of the output of ten randomly-generated CPPN pairs, each defining the key musical relationships between the scaffold and output accompaniment. These accompaniments can be viewed in a graphical depiction (as shown in the screenshot in figure 3) or in standard musical notation. They can be played and heard through either MIDI or MP3 formats. The user-guided process of exploration that combines and mutates these candidates is called *interactive evolutionary computation* (IEC) (Takagi, 2001). Because each accompaniment is encoded by two CPPNs, evolution can alter both the pitch and rhythmic CPPNs or adjust them individually.

The user combines and explores accompaniments in this space by selecting and rating one or more accompaniments from one generation to parent the individuals of the next generation. The idea is that the good musical ideas from both the rhythmic and pitch functions are preserved with slight alterations or combined to create a variety of new but related functions, some of which may be more appealing than their parents. The space can also be explored without combination by selecting only a single accompaniment. The next generation then contains slight mutations of the original functions.

While IEC inherently facilitates these types of creativity, the approach in this paper extends the reach of transformational creativity offered by FSMC. Previously, FSMC generated single-voice accompaniments to be played with a fully-composed, preexisting human

piece (Hoover, Szerlip, and Stanley, 2011a,b). This paper introduces a new layering technique whereby generated accompaniment from previous generations can serve as inputs to new CPPNs that then generate more layers of harmony. The result is the ability to spawn an entire multi-layered piece from a single monophonic starting melody.

One such layering approach is performed by generating one new monophonic accompaniment at a time. The first layer is the monophonic melody composed by the human user. The second layer is generated through FSMC from the first. The third layer is then generated through FSMC by now inputting into the CPPNs the first *and* second layers, and so on. All of the layers are finally combined to create an entire accompaniment, resulting in accompaniments that are functionally related to both the initial melody and previous accompaniment lines. In this way, each accompaniment line is slightly more removed from the original melody and subsequent accompaniment lines are based functionally on both the scaffold and previously-generated lines.

To create accompaniments more closely related to the original melody, another layering technique is for users to generate *all* accompaniment layers from only the single monophonic starting point. For this purpose, the CPPNs are given enough outputs to represent all the instruments in the accompaniment at the same time. Because the melody and the accompaniments are functionally related, any accompaniment will follow the contours of the melodic starting point. However, in this case, the only influence on each accompaniment is this starting point itself, yielding a subtly different feel.

With either of these approaches or a combination of them users can further influence their accompaniments by holding constant the rhythm CPPN or pitch CPPN while letting the other evolve. Interestingly, when two accompaniments share the same rhythm network but differ in the pitch network slightly, the two monophonic instruments effectively combine to create the sound of a polyphonic instrument. Similarly, the pitch networks can be shared while the rhythm networks are evolved separately, creating a different sound. Notice that this approach requires no musical expertise to generate multiple lines of accompaniment.

Experiments

The experiments in this paper are designed to show how users can generate multipart pieces from a single monophonic melody with FSMC. They are divided into accompaniment generation and a listener study that establishes the quality of the compositions.

Accompaniment Generation

For this experiment, three members of our team composed in total three monophonic melodies. From each of these user-composed melodies, a multipart accompaniment was generated through FSMC by the author of the originating melody. Two other multipart accompaniments were generated for the folk song, *Early One*

Morning. We chose to include each of these FSMC composers, who were undergraduate independent study students at the University of Central Florida, as authors of this paper to recognize their pioneering efforts with a new medium. The most important point is that no musical expertise need be applied to the final creations beyond that necessary to compose the initial monophonic melody in MIDI format. Thus, although results may sound consciously arranged it is important to bear in mind that all the polyphony you hear is entirely the output of FSMC. The original melodies, accompaniments, and CPPNs are available at <http://eplex.cs.ucf.edu/fsmc/iccc2012>. The program, called MaestroGenesis, is available at <http://maestrogenesis.org>.

As noted in the approach, FSMC provides significant freedom to the user in how to accumulate the layers of a multipart piece. In general, the user has the ability to decide from which parts to generate other parts. For example, from the original melody, five additional parts could be generated at once. Or, instead, the user might accumulate layers incrementally, feeding each new part into a new CPPN to evolve yet another layer. Some layers might depend on one previous layer, while other might depend on multiple previous layers. In effect, such decisions shape the subtle structural relationships and hence aesthetic of the final composition. For example, evolving all of the new parts from just the melody gives the melody a commanding influence over all the accompaniment, while incrementally training each layer from the last induces a more delicate and complex set of harmonious partnerships. As the remainder of this section describes, the student composers took advantage of this latitude in a variety of ways

Early One Morning, (Song 1) versions 1 and 2 with four- and five-part accompaniments began from an initial monophonic melody transcribed from the traditional, human-composed folk song. The second layer is identical in both versions and was evolved from Early One Morning itself. The third, fourth, and fifth parts of version 1 were all evolved from the second layer. The third, fourth, fifth, and sixth parts of version 2 were evolved from the pitch network of the second layer of version 1, and the rhythm network from the original Early One Morning monophonic melody. This experiment illustrates that the results with FSMC given the same starting melody are not deterministic and in fact do provide creative latitude to the user even without the need for traditional composition techniques.

Song 2 started from an original monophonic melody composed by undergraduate Marie E. Norton. The second layer was added by inputting this melody into the rhythm and pitch networks of the subsequent accompaniment populations. This second layer then served as input to the pitch and rhythmic CPPNs for layers 3 and 4. The pitch CPPN for layer 5 consisted of layer 2, but the rhythm network only had a bias input. Finally, the inputs for the pitch network for layer 6 were layers 3, 4, and 5, while the inputs to the rhythm CPPN were

layer 4 and a measure timing signal first introduced for FSMC by Hoover and Stanley (2009) that gives the network a sense of where the song is within the measure. All of the layers finally combined to create a single, multipart piece in which each line is functionally related to the others. Each layer took as few as three to as many as five generations to evolve.

For Song 3, Zachary Merritt first created a layer that influences most of the other layers, but is not heard in the final track. The fourth layer was generated from the third, which is influenced by the monophonic melody and the unheard layer. The fifth layer was generated from the population of the fourth layer with the rhythm network held constant to create a chordal feel. The sixth layer was generated from only the initial starting melody and a special timing signal that imparts a sense of the position in the overall piece (Hoover and Stanley, 2009). Similarly, the seventh layer is generated from only the initial starting melody, but adds a separate function input, $\sin(\pi x)$, where x is the time in measure. Although there are seven layers described in this experiment, only six were selected to be heard, meaning that there is a five-part accompaniment.

Trevor A. Brindle created an initial piece and evolved all five accompaniment lines for Song 4 directly from it. Instead of inputting results from previous generations, he started new runs for each voice from the same scaffold, giving a strong influence to the melody.

Notice that the key decisions made by the users are in general from which tracks to generate more tracks. Of course the users also performed the IEC selection operations to breed each new layer. Importantly, none such decisions require musical expertise.

Listener Study

The contribution of users to the quality of the generated works and accordingly the effectiveness of the creativity enhancement is evaluated through a listener study. The study consists of five surveys, one for each generated arrangement. The surveys present two MP3s to the listener, who is asked to rate the quality of both. The first MP3, called the *collaborative accompaniment*, is an arrangement resulting from the collaboration of the author with the program (i.e. the two versions from Early One Morning or Songs 2, 3, or 4). The second, called the *FSMC-alone accompaniment*, is generated by the program *alone*. That is, a random pitch CPPN and a random rhythm CPPN are provided the same monophonic starting melody as the collaborative accompaniment and their output is taken as the FSMC-alone accompaniment. Thus the factor that is isolated is the involvement of the *human user*, who is not involved in the FSMC-alone accompaniment. However, it is important to note that the FSMC-alone accompaniments do not actually *sound* random because even if the CPPNs are generated randomly, they are still functions of the same scaffold, which tends even in the random case to yield outputs that sound at least coherent (which is the motivation for FSMC in the first place). Thus this

study investigates whether the human user is really able to make a creative contribution by leveraging FSMC.

A total of 129 students participated in the study. The full survey is available at <http://eplex.cs.ucf.edu/fsmc/icc2012/survey>, but note that in the administered surveys, the order of the MP3s was random to avoid any bias. The users were asked to rate each piece with the following question:

Rate MIDI i on a scale of one to ten. (1 is the worst and 10 is the best),

where i refers to one of the ten generated works. The idea is that if the user-created arrangements are rated higher than those generated by FSMC-alone, the user's own input likely positively influenced the outcome. While this study focuses on the quality of output, the degree to which FSMC *enhances* creativity will be addressed in future work.

Results

The generated accompaniments and original scaffold discussed in this section can be heard at <http://eplex.cs.ucf.edu/fsmc/icc2012>.

Accompaniments

Samples of the scores for the two arrangements created to accompany Early One Morning are shown in figure 4. The layers are shown in order from top to bottom in both versions (layer 1 is the original melody). Layer 2, which is the same in both versions, is heard as violin II in version 1 and viola in version 2.

An important observation is that the violoncello part in version 1 follows the rhythm of the initial starting melody very closely while the pitch contour differs only slightly. While the viola and double-bass parts differ in both pitch and rhythm over the course of the song, both end phrases and subphrases on the tonic note, F, in many places over the course of the piece, including measure 4 in figure 4a. Version 2, on the other hand, contains many rhythmic similarities (i.e. the eighth note patterns contained in the keyboard I, viola, keyboard II, and the violin II parts), but illustrates distinct pitch contours. Together, the two versions illustrate how a single user can generate different accompaniment from the same initial monophonic starting melody and how the initial melody exerts its influence both rhythmically and harmonically.

Songs 2, 3, and 4 exhibit a similar effect: rhythmic and harmonic influence from the original melody, yet distinctive and original accompaniment nevertheless. The result is that the overall arrangements sound *composed* even though they are evolved through a breeding process. The next section provides evidence that impartial listeners also appreciate the contribution of the human user.

Violin I (Layer 1)
Violin II (Layer 2)
Violoncello (Layer 3)
Double Bass (Layer 4)
Viola (Layer 5)

(a) Early One Morning Version 1

Keyboard I (Layer 1)
Viola (Layer 2)
Keyboard II (Layer 3)
Electric Bass (Layer 4)
Violin I (Layer 5)
Violin II (Layer 6)

(b) Early One Morning Version 2

Figure 4: **Early One Morning**. The first four measures of versions 1 and 2 of Early One Morning illustrate how a single user with the same monophonic starting melody can direct the accompaniment in two different ways that nevertheless both relate to the initial melody. Because the accompaniments share two of their layers, they sound related. However, through timbre selection and the evolution of two and three distinct layers in versions 1 and 2 respectively, the user imparts a different feel.

Listener Study Results

The results of the listener study in figure 5 indicate that all of the collaborative accompaniments are rated higher than those generated with FSMC alone, with three out of five (Song 1 version 2, Song 4, and Song 5) displaying significant difference ($p < 0.05$; Student's paired t-test). Taken all together, the collaborative accompaniments sound highly significantly more appealing than those generated with FSMC alone ($p < 0.001$; Student's paired t-test). These results indicate that not only does FSMC provide a structurally plausible search space, but that it is possible to explore such a space without applying musical expertise. That is, the results suggest that the user input significantly improves the perceived quality of the generated compositions.

Discussion

A key feature of figure 4 is that the collaborative accompaniments generated by users with the assistance of FSMC follow the melodic and rhythmic contours of the original scaffold. Furthermore, the listener study suggests that FSMC helps the user establish and explore musical search spaces that may otherwise have been inaccessible.

While the users search this space through IEC, which facilitates the combination of musical ideas and the exploration of the space itself, an interesting property of this search space is its robustness; even FSMC-alone accompaniments, which are created without the benefit of human, subjective evaluation, can sound plausible. However, when coupled with the human user, this approach in effect *transforms* the user's own internal search space of possible accompaniments to one constrained by functional scaffolding.

While the quantitative data suggests the merit of collaborative accompaniments, music is inher-

ently subjective. Therefore, it is important for the reader to judge the results for his or herself at <http://eplex.cs.ucf.edu/fsmc/iccc2012> to fully appreciate the potential of the FSMC method.

One interesting direction for future work is to explore new interpretations for the output of the pitch functions. Currently, accompaniment pitches are interpreted as discrete note values, a process that limits the instrument to playing the same note each time a given combination of notes occurs in the scaffold. However, by interpreting the output as a change in pitch (i.e. horizontal interval) rather than an absolute pitch, instruments can select any note to correspond to a particular combination depending on where in the piece it is occurring. In this way, an even larger space of musical possibilities could be created.

Perhaps most importantly, with only a single, monophonic melody, users could compose entire multipart pieces without the need for musical expertise. Even if not at the master level, such a capability opens to the novice an entirely new realm of exploration.

Conclusion

This paper presented an extension to functional scaffolding for musical composition (FSMC) that facilitates a human user's creativity by generating polyphonic compositions from a single, human-composed monophonic starting track. The technique enables creative exploration by helping the user construct and then navigate a search space of candidate accompaniments through a breeding process akin to animal breeding called interactive evolutionary computation (IEC). These collaborative accompaniments bred by users were judged by listeners against those composed only through FSMC. Overall, listeners liked collaborative accompaniments more than the FSMC-alone accompaniments. Most importantly, a promising poten-

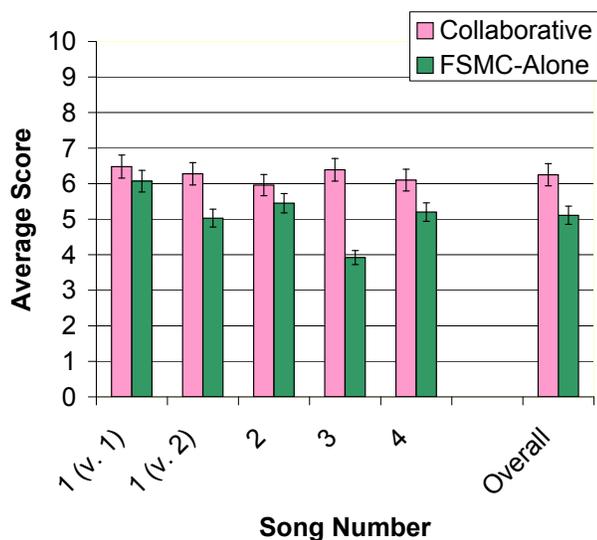


Figure 5: **Listener Study Results.** The average rating (by 129 participants) from one to ten of both the collaborative and FSMC-alone accompaniments are shown side-by-side with the lines indicating a 5% error bound. The overall results for the listener study indicate that on average the collaborative accompaniments are of significantly higher perceived quality than FSMC-alone.

tial for creativity enhancement in AI is to open up the world of the amateur to the domain once only accessible to the expert. The approach in this paper is a step in this direction.

Acknowledgements

This work was supported in part by the National Science Foundation under grant no. IIS-1002507 and also by a NSF Graduate Research Fellowship.

References

Ando, D., and Iba, H. 2007. Interactive composition aid system by means of tree representation of musical phrase. In *IEEE Congress on Evolutionary Computation (CEC)*, 4258–4265. IEEE.

Biles, J. 1998. Interactive GenJam: Integrating real-time performance with a genetic algorithm. In *Int. Computer Music Conf. (ICMC 98)*, 232–235.

Boden, M. A. 2004. *The Creative Mind: Myths and Mechanisms*. Routledge, second edition.

Boden, M. A. 2007. *Creativity and Conceptual Art*. Oxford:Oxford 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. 1987. An expert system for computer-assisted composition. *Computer Music Journal* 11(4):30–46.

Cybenko, G. 1989. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems (MCSS)* 2(4):303–314.

Holtzman, S. R. 1980. A generative grammar definition language for music. *Interface* 9(1):1–48.

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.; 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.

Jacob, B. L. 1995. Composing with genetic algorithms. In *Proc. of the 1995 International Computer Music Conference*, 425–455. Intl. Computer Music Association.

Keller, R. M.; Morrison, D.; Jones, S.; Thom, B.; and Wolin, A. 2006. A computational framework for enhancing jazz creativity. In *Proceedings of the Third Workshop on Computational Creativity, ECAI 2006*.

Kippen, J., and Bel, B. 1992. *Modeling Music with Grammars: Formal Language Representation in the Bol Processor*. Academic Press London. 207–238.

Marsden, A. 2000. *Readings in Music and Artificial Intelligence*. Harwood Academic Publishers. chapter Music, Intelligence, and Artificiality, 18.

McCormack, J. 1996. Grammar based music composition. *Complex Systems* 96:321–336.

Ralley, D. 1995. Genetic algorithms as a tool for melodic development. In *Proc. of the 1995 Intl. Computer Music Conf.*, 501–502. Intl. Computer Music Assoc.

Roads, C. 1979. Grammars as representations for music. *Computer Music Journal* 3(1):48–55.

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.

Stanley, K. O., and Miikkulainen, R. 2002. Evolving neural networks through augmenting topologies. *Evolutionary Computation* 10:99–127.

Stanley, K. O. 2007. Compositional pattern producing networks: A novel abstraction of development. *Genetic Programming and Evolvable Machines Special Issue on Developmental Systems* 8(2):131–162.

Takagi, H. 2001. Interactive evolutionary computation: fusion of the capabilities of ec optimization and human evaluation. *Proceedings of the IEEE* 89(9):1275–1296.

Todd, P. M., and Werner, G. M. 1999. Frankensteinian methods for evolutionary music. *Musical Networks: Parallel Distributed Perception and Performance* 313–340.

Zicarelli, D. 1987. M and jam factory. *Computer Music Journal* 11(4):13–29.

Zicarelli, D. 2002. How I learned to love a program that does nothing. *Computer Music Journal* 26(4):44–51.