Interactively Evolving Harmonies through Functional Scaffolding

In: *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2011)*. New York, NY: ACM
Winner of the Best Paper Award in the Digital Entertainment Technologies and Arts (DETA) Track

Amy K. Hoover
University of Central Florida
Department of EECS
Orlando, FL, 32816
ahoover@eecs.ucf.edu

Paul A. Szerlip
University of Central Florida
Department of EECS
Orlando, FL, 32816
paul.szerlip@gmail.com

Kenneth O. Stanley
University of Central Florida
Department of EECS
Orlando, FL 32816
kstanley@eecs.ucf.edu

ABSTRACT

While the real-time focus of today's automated accompaniment generators can benefit instrumentalists and vocalists in their practice, improvisation, or performance, an opportunity remains specifically to assist novice composers. This paper introduces a novel such approach based on evolutionary computation called *functional scaffolding for musical composition* (FSMC), which helps the user explore potential accompaniments for existing musical pieces, or *scaffolds*. The key idea is to produce accompaniment as a *function* of the scaffold, thereby inheriting from its inherent style and texture. To implement this idea, accompaniments are represented by a special type of neural network called a compositional pattern producing network (CPPN), which produces harmonies by elaborating on and exploiting regularities in pitches and rhythms found in the scaffold. This paper focuses on how inexperienced composers can personalize accompaniments by first choosing any MIDI scaffold, then selecting which parts (e.g. the piano, guitar, or bass guitar) the CPPN can hear, and finally customizing and refining the computer-generated accompaniment through an interactive process of selection and mutation of CPPNs called interactive evolutionary computation (IEC). The potential of this approach is demonstrated by following the evolution of a specific accompaniment and studying whether listeners appreciate the results.

Categories and Subject Descriptors: J.5 [**Arts and Humanities**]: fine arts, music

General Terms: Algorithms

Keywords: Accompaniment Generation, Interactive Evolutionary Computation (IEC), Compositional Pattern Producing Networks (CPPNs), Music, NEAT

1. INTRODUCTION

An exciting application of computer technology to music in recent years is accompaniment generation. Most such systems compose in real-time by listening to a performance, computing a musical response, and performing with the musician. These systems learn or deduce musical structure online or offline through probabilistic models, grammars, and expert knowledge [10, 16, 23, 1]. Musical decisions, i.e. what to play and when, are thus ultimately made by the programmers and algorithms. While such systems help performers practice, improvise, and entertain, they focus less on assisting novice composers to enhance their own creativity, which is the topic of this paper.

Instead of forcing computers to make subjective decisions, the proposed approach implements *interactive evolutionary computation* (IEC), a process that lets the human, rather than the computer, appraise accompaniments. In particular, a set of candidate accompaniments are presented to the user, who rates each piece in the set. The next set is then generated based on the preferred individuals of the previous generation. That way, IEC turns the combination of the user and computer into a *team* that collaborate to incrementally enhance the accompaniment intuitively. However, the success of IEC depends on the accompaniment's representation; when the IEC search space of candidates is desolate from poor representation, users quickly fatigue [22].

For that purpose, this paper introduces a method called *functional scaffolding for musical composition* (FSMC) that builds on an existing composition, or *scaffold*, to create natural-sounding accompaniments evolved through IEC. FSMC significantly extends a previous theory of accompaniment generation by Hoover and Stanley [8], which focused exclusively on percussion accompaniment, by adding the ability to generate harmonization. The main insight behind FSMC is that different instrumental parts in musical compositions are *functionally related* to each other. Therefore, if some parts are already given, then new parts (i.e. the accompaniment) can be generated as a *function* of the original parts.

In particular, scaffolds are input to a type of artificial neural network called a compositional pattern producing network (CPPN), which computes a function that transforms the scaffold. The CPPN in effect manipulates the original pitch and rhythmic patterns through function composition [19] (which produces transformations) to generate realistic accompaniment. Because CPPNs represent music as patterns, i.e. as opposed to representing each note separately, most outputs inherit at least some appealing qualities from the scaffold. FSMC users thus can search accompaniment space effectively without experiencing too much fatigue. Because of the bias provided by FSMC, at every iteration of this algorithm, several viable accompaniment options are

GECCO'11, July 12–16, 2011, Dublin, Ireland.

available to the user, which means the IEC process focuses on principled refinement rather than on escaping chaos to discover order.

To establish the potential of FSMC to aid novice composers (and perhaps to inspire ideas for experts as well) in this paper, the evolution of a particular accompaniment is closely followed and analyzed, showing how the user and computer collaborate to produce an effective process of refinement without the need for expert musical knowledge. The results of this progression are also validated through a listener study that asks independent participants to rate different steps in the process. The main result is that FSMC-based IEC could be a step towards new tools for expanding the pleasure of creating multipart pieces to a wider population than is currently possible. In effect, such a tool can provide a stepping stone that inspires users to learn more about music who otherwise might not. In addition, by casting IEC as a metaphor for creative exploration in general, it shines a light on how creative endeavors and search can interact to yield meaningful discoveries.

2. BACKGROUND

Many approaches to automating accompaniment leave musical decisions entirely to the computer. For example, ImprovGenerator creates drum pattern accompaniment from a live drumming performance. The patterns are generated by context-free grammars and transition probability models that input live patterns, infer grammatical rules, and apply them probabilistically [10]. This method incorporates the performer's musical sensibilities, but the performer has little input into the accompaniment generated. Similarly, Santarosa et al. [16] approach accompaniment generation with a two-tiered genetic algorithm, evolving both accompaniments and the fitness functions with which they are rated. In both approaches, the human user can only control the output insofar as they can manipulate the input stream. Because important decisions on the accompaniment itself are made by the algorithm, such approaches are best suited to live performances as opposed to novice composers aiming to explore ideas for new accompaniments.

In contrast, some approaches incorporate humans through interactive evolutionary computation (IEC [22]) to address the inherent subjectivity in musical judgments [9, 13, 3, 4, 8, 24]. The idea is that humans can rate candidate accompaniments rather than an explicit fitness function. IEC originated in Richard Dawkins' book, *The Blind Watchmaker*, in which he described a simple program called *Biomorphs* that is meant to illustrate evolutionary principles [5]. The program displays a set of several pictures (called Biomorphs) on the screen at one time. The user then selects from among those pictures (called the *population*) his or her favorite. From that selection, a new generation of *offspring* is spawned that replace the original population. Because the offspring are generated through slight mutations of the underlying genes of the selected parents, they tend to resemble their parents while still suggesting novel traits. In this way, over many generations, the user in effect *breeds* new forms.

IEC can encompass a variety of digital media [15, 22], including images [11, 26, 6], movies [25], three dimensional models [14], and music [9, 13, 3, 4, 8, 24]. GenJam is an example IEC application in which users help to build improvisational accompaniments or soloists by rating the measures and phrases as either good or bad [3]. Yet while users can interact with these systems, the size of the search space remains an obstacle to novice exploration. Thus the unexploited opportunity at the focus of this paper is to borrow from the creative seed *already* in the scaffold to enhance that creativity further.

The inspiration for this idea is a prior approach called NEAT Drummer, which harnesses creativity present in an existing piece (i.e. scaffold) to generate percussion accompaniment [8, 7]. Each rhythm in NEAT Drummer is represented by a CPPN [19] and evolved interactively through the NeuroEvolution of Augmenting Topologies (NEAT) method [20, 21]. NEAT is a method for evolving networks that sometimes adds new nodes or connections through mutations. In NEAT Drummer, the CPPN representation and scaffold constrain the search space to promising candidates. Furthermore, the relationship between scaffold inputs and drum pattern accompaniments can complexify over time through NEAT, but the survival of structural additions and deletions depend on the choices made through IEC. That way, users guide the creative explorations by both choosing the input to the system, i.e. from which input tracks to generate accompaniment, and by deciding which accompaniments sound plausible.

The FSMC method introduced in this paper significantly extends the ideas in NEAT Drummer. Whereas NEAT Drummer can only generate percussion and rhythm, FSMC generates entire harmonies through a new CPPN representation, explained next.

3. APPROACH

Extending the idea in NEAT Drummer, harmonies in functional scaffolding for musical composition (FSMC) are generated from existing compositions. These compositions form a *scaffold* from which accompaniments are built. However, unlike in NEAT Drummer, these scaffolds include timing information *and* pitch information, thereby providing the foundation for harmonization.

Because music is essentially repeating or partially-repeating patterns over time, musical parts can be conceived as functions of time. Following this idea, FSMC represents accompaniments as a *function* that transforms pitches and rhythms from input tracks (i.e. the scaffold) into a pattern interpreted as the accompaniment. In particular, this function is encoded in FSMC by a compositional pattern producing network (CPPN), whose representation is detailed in the next section. Outputs from CPPNs are interpreted as accompaniments that thereby closely follow contours of the original song. Users then interactively explore the search space of such functions to personalize accompaniments through IEC.

This section details the CPPN representation of the function that generates the accompaniment and the interactive process through which the user searches the space of such functions.

3.1 Representing Relationships

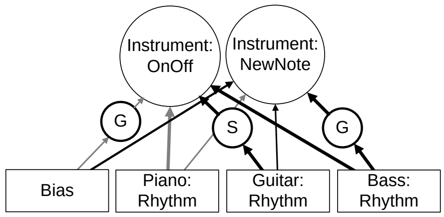
A crucial aspect of FSMC accompaniment is its representation, which is a CPPN. The main difference between CPPNs and neural networks (in which every node usually computes the same activation function) is that hidden nodes in the same CPPN can compute a diversity of activation functions [19]. These include Gaussian, sine, and sigmoid functions. This representation is well-suited to music because such functions compactly encode useful regularities such as symmetry (through the Gaussian) and repetition (through the sine). Because CPPNs can usually evolve arbitrary topologies through the NEAT method, such functions often end up composed with each other, which leads to expressing increasingly complex nonlinear relationships, such as repetition with variation (e.g. by composing sine and sigmoid functions). The pattern-generating capability of CPPNs has been demonstrated also in image generation on the Picbreeder online service [18, 17]. In effect, in the present work, the CPPN embodies the function that transforms the scaffold into its own accompaniment.

In general, the inputs to CPPNs in FSMC are instrument channels of the existing composition and the outputs control different aspects of the instruments in the accompaniment. It is important to note that input musical data must be *sequenced* (i.e. not raw audio), such as notes in the MIDI format. That way, the lengths and pitches of notes in the scaffold are known. The harmonic accompaniment (e.g. clarinet or trumpet) is represented by two CPPNs: one controls the rhythm and the other dictates pitch. Figure 1 shows an example pair of CPPNs for rhythm and pitch generation. In figure 1a, inputs (shown at bottom) are rhythms from the piano, guitar, and bass parts of an existing song. One output is a volume node called OnOff, which indicates whether a note or rest is played and the volume at which the note is heard. The other rhythm output, called NewNote, determines whether notes are rearticulated. Figure 1b illustrates the pitch network paired with the given rhythm network. The output of this network decides which pitch the accompaniment track plays if a note is sounded, based on the pitches of every instrument in the scaffold.

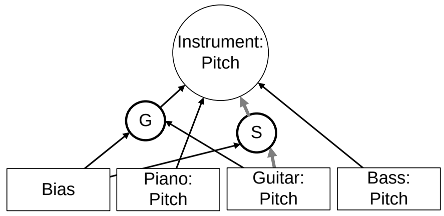
To understand how the networks in figure 1 generate accompaniment, note that the notes and timing information that are fed into the CPPNs' inputs are provided sequentially over time over a series of ticks (usually spaced at sixteenth-note durations, depending on the duration of the shortest notes). That is, the CPPNs in effect "listen" to the piece in order and thereby output accompaniment that is also in correct temporal order. Specifically, FSMC inputs track *rhythms* to the rhythm CPPN through the spike and decay method first introduced in NEAT Drummer [8, 7]: When a note attacks in the input track, a maximum value is sent to the network that linearly decays until the note ends. That way, the CPPN knows *where* within each note duration it is at every tick (which effectively provides timing information). On the other hand, track *pitches* are input as pitch classes (e.g. both C4 and C5, which are an octave apart, are input as a C). The output pitches (which form the accompaniment) are then normalized to the key of the existing composition. These pitches can span several octaves, depending on the particular instrument selected for output. The input representations just described for rhythm and pitch are depicted in figure 2.

It is important to note that the rhythm and pitch CPPNs are separated intentionally because combining them into a single CPPN would in effect imply that *times* within a piece are semantically similar to *pitches* of notes. Such a confusing conflation leads to a jumble of incoherent patterns, as preliminary experiments with such a setup confirmed. Thus the separation is an important aspect of the representation.

The overall effect is that the CPPNs output a temporal pattern that is functionally related to the scaffold in both rhythm and pitch. This relationship is what makes the accompaniment sound sensible. However, the particular transformation of the scaffold computed by the CPPN is *evolved*, which means it can be arbitrarily complex and nuanced. The next section explains how the user influences this evolution.



(a) Rhythm CPPN



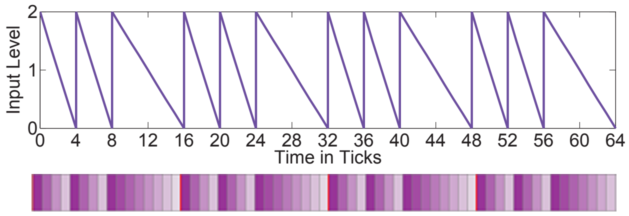
(b) Pitch CPPN

Figure 1: Example FSMC CPPN Pair. Together, the CPPNs shown in (a) and (b) generate pitched accompaniment. The rhythm network in (a) indicates volume or resting through the OnOff node and also whether or not to rearticulate the note. The pitch network in (b) controls which pitch is played if a rest is not indicated. Hidden nodes (e.g. sigmoids and Gaussians), which would have evolved through NEAT, are visible within the network topologies. In this way, the accompaniment, which is the output of the CPPN, is a *function* of the existing scaffold, thereby borrowing from its structure.

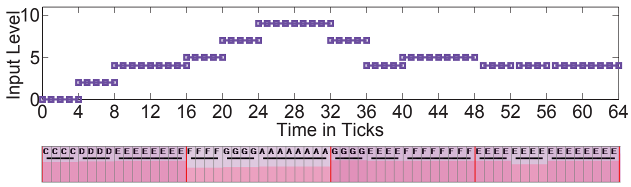
3.2 Choosing Scaffolds and Evolving Harmonies

Users select accompaniments through IEC. After choosing initial network configurations, i.e. inputs for the rhythm and pitch networks and instrument outputs, FSMC presents users with a population of possible accompaniments to the scaffold (figure 3). These candidates are then rated by the user. The user rates any number of pieces in the current population, all of which influence the next generation of accompaniments. If mutation rates are high, the character between generations varies greatly. Otherwise, with low mutation rates, the user evolves pieces with similar character. Mutations in CPPNs generally modify connection weights or add or remove connections and nodes, following the NEAT method [20].

While IEC has previously been applied to music generation [13, 12, 2, 3], the hope in FSMC is that the CPPN representation of the functional relationship between scaffold and accompaniment will allow a holistic evolution of song-wide



(a) Rhythm Inputs



(b) Pitch Inputs

Figure 2: CPPN Input Representation. The spike-decay representation for rhythms is shown in (a) and the pitch representation is shown in (b). Both such inputs are depicted in two ways: The first is a continuous-time graph that shows decaying spikes for timing and the pitch level for pitch. The second is a discrete-time representation of what is actually input into the network at each discrete time step, which is represented by darkness for rhythm and height for pitch. Because the network samples time discretely, results in this paper are also depicted in the discrete-time format. In this way, this figure gives a sense of exactly what the CPPN "hears" (for each instrument in the scaffold) as it generates accompaniment.

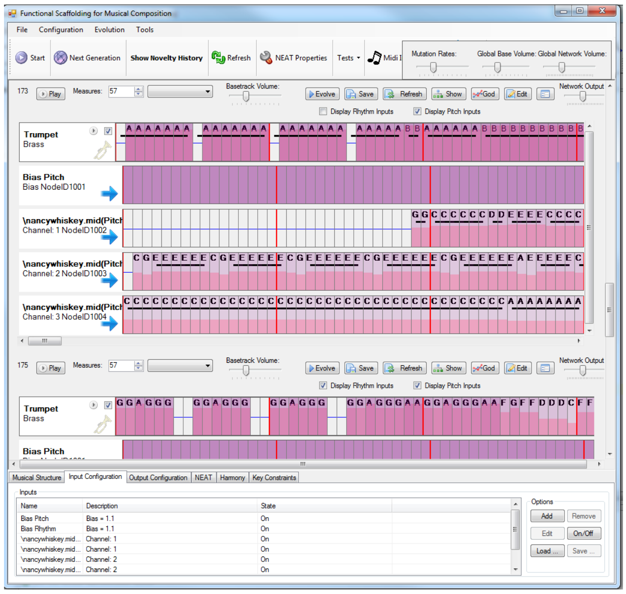


Figure 3: FSMC Graphical User Interface. Accompaniments in FSMC are presented both visually and sonically. Unlike image evolution, in which users can quickly evaluate the population [22], listening to MIDIs takes time. The visual representations help users decide which MIDIs are worth this extra time, thus speeding up evolution. There are ten individuals in a population, which are all displayed at once. The user rates individuals by clicking on ratings through a drop-down menu.

patterns. Instead of manipulating single notes or features of a composition, FSMC evolves entire functional relationships, thereby ensuring that the search space at least only considers accompaniments with some relationship to the scaffold.

Selecting the scaffold is itself an important task. It requires choosing to which instrument tracks the rhythm and pitch networks should listen. While beginners can easily choose appropriate inputs to create appealing accompaniments, refined choices can significantly influence the piece. For example, whether or not the rhythm network listens to a fast-changing instrument can impact the complexity of the corresponding output accompaniment. In fact, chosen tracks do not have to be the same for each network (e.g. the rhythm network can have a piano and guitar input while the pitch network only has a bass guitar input).

Because the parts of the scaffold themselves are human-composed and thereby sound appealing, accompaniment built from any combination of such tracks ends up following the contours of the original song. However, depending on the specific inputs selected and the internal network structure, the relationship between selected inputs and the accompaniment may be of varying complexity. NEAT [20] (the underlying evolutionary algorithm that evolves CPPNs), which occasionally adds new structure to CPPNs (and can also remove it in the implementation in this paper) allows such complexity to increase or decrease following the preferences of the IEC user.

4. EXPERIMENTS

The main focus of this paper is on the effect of the IEC process on discovering high-quality accompaniment. The hope is that IEC coupled with FSMC leads to definitive improvement that requires no musical expertise and only minimal effort (i.e. not many generations), thereby avoiding the fatigue effect that plagues many experiments in IEC [22]. This section describes the IEC experiment and the listener study designed to illuminate its results.

4.1 Evolving Accompaniment

The experiment focuses on accompaniment evolved by the authors to the folk song Bad Girl's Lament, which was originally arranged in MIDI format by Barry Taylor and is redistributed with his permission. An accompaniment generated for Scarborough Fair, also arranged by Barry Taylor, is also presented to give insight into some of the challenges faced by FSMC. While FSMC ultimately will be applied to incomplete compositions by amateur musicians, by focusing in this first study on evolving accompaniment to a well-regarded *existing* song, it is possible to begin with a baseline level of quality. That way, we can ask whether accompaniments maintain or degrade the quality of the initial song (which, after all, is complete even without additional accompaniment).

By functionally relating the accompaniment to the scaffold, the hope is that it is possible to quickly evolve into an area of the search space that rivals the quality of the initial composition so that the search effectively becomes a search for high-quality variety rather than a struggle to reconstitute lost quality.

For Bad Girl's Lament, an evolutionary progression between generations 1 and 12 is studied by highlighting important milestones at generations 1, 6, and 12. This 12-generation progression took about thirty minutes in total for the user to complete; most of the time was spent listening to candidate accompaniments. Inputs to the rhythm and pitch CPPNs are the piano and harpsichord channels from the scaffold.

While the particular run of Bad Girl's Lament chosen for analysis is anecdotal, most results with other pieces exhibit similar features and dynamics. Thus the hope in this paper is to provide deeper insight into what exactly FSMC does when combined with IEC by focusing on the details of a particular progression in the limited space of a conference paper. In addition, an example of a problem that occurs with some songs is demonstrated with Scarborough Fair. Later reports will identify more general elements of FSMC, but without the level of individual detail possible here.

Accompaniments are evolved with a CPPN mutation rate and crossover rate of 0.3. The NewNote threshold is also 0.3. Furthermore, when the OnOff output in the rhythm network (which also indicates volume) falls below 0.3, no note is played. Population size was 10 per generation and fitness is scored as 0, 1, or 2, where 0 indicates an unappealing track and 2 denotes the best tracks.

4.2 Listener Study

To understand the effect of evolution on subjective appreciation, a total of 60 listeners, all of whom are students in a diversity of majors at the University of Central Florida, participated in a survey after listening to the evolved variants of Bad Girl's Lament. In particular, without knowing which is which, they listened to (1) an intentionally poor-quality control with inappropriate accompaniment (which helps to establish that participants indeed generally agree on something subjective), (2) the original Bad Girl's Lament without accompaniment, (3) the accompaniment selected from the first generation of IEC, (4) the accompaniment selected from the sixth generation of IEC, and (5) the final selected accompaniment from generation 12. For each of these variants, the listener was asked:

> 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 five variants, which are available for listening online at
`http://eplex.cs.ucf.edu/fsmc/gecco2011/`.

By establishing the perceived quality of a respected composition, it becomes possible to estimate how well evolution can maintain that professional standard even though FSMC with IEC incorporates *no prior musical knowledge or expertise*. In fact, the important deeper aim in this experiment is to suggest that through FSMC, evolutionary-assisted music generation can reach a high level of quality even with almost no musical theory whatsoever. Of course, given such a result, a serious application of the technology could then augment the FSMC core with more refined knowledge.

5. RESULTS

To appreciate the results in this section it is important to experience the generated tracks. Thus all the selections discussed in this section can be heard at `http://eplex.cs.ucf.edu/fsmc/gecco2011/`. The section begins by analyzing the evolutionary progress of accompaniments to Bad Girl's Lament and then turns to the listener study.

5.1 Accompaniments

Results in this section are reported through figures that are designed to demonstrate the relationship between the CPPN inputs and outputs as the song progresses over time. Boxes, or ticks of the CPPN, are read from left to right and show what each instrument is playing at a particular instance of time. Darker box shading denotes volume and the shading height denotes the relationship between pitches in the accompaniment; higher pitches have taller shadings. The absolute pitch (i.e. note) is written in bold at the top of each rectangle. A slightly thicker dividing line between columns denotes a measure break.

At each tick, a pitched output can either sustain the note from the previous tick, play a new note, or rest. Sustains are represented by a thick line struck through the box and rests are indicated by a similar but thinner (and slightly lower) line over a white box. Both rhythm and pitch inputs to the CPPN are also shown in visualizations. It is important to note that rhythm inputs represent the special spike-decay format introduced in figure 2a while pitch inputs are simply pitch levels (as in figure 2b).

Figure 4 shows measures 1, 2, 13, and 14 of generations 1, 6, and 12 of Bad Girl's Lament with evolved accompaniments. The introduction in measures 1 and 2 of the first generation is particularly unnatural; not only are notes rearticulated on each tick, lending a rushed and choppy feel to the piece, but only one pitch is played. On the other hand, the introductory measures in generations 6 and 12 are smoother (i.e. with sustained notes) and conform to the rhythmic and pitch contours of the original song. They alternate between the notes B and D, whereas generation 1 only plays D. The rhythm becomes more sophisticated across generations as well. Both generations 6 and 12 have significantly less choppy introductions than the first, but the rhythm CPPN in generation 12 creates a swing-style pattern.

The pitches in measures 13 and 14 of the first generation differ significantly from those created for generations 6 and 12. Pitches in generation 1 ascend across notes A, B, and C#, followed by B, C#, and D in the next measure. However, in generations 6 and 12, the pattern more closely follows the harpsichord input from the scaffold, demonstrating the influence of the functional relationship on the evolved progressions. For example, they all travel from B to D and back to B in the first measure. However, in the second measure, generation 12 falls back to a C# rather than the B selected for generation 6. This variation imparts a progressive resolution that is missing in the thirteenth and fourteenth measures of generations 1 and 6.

While the three depicted generations in Bad Girl's Lament exhibit some similar characteristics (e.g. each individual plays more rests when the piano from the scaffold is resting early in the composition), they progressively change over evolutionary time. For example, while generations 6 and 12 are rhythmically similar, generation 12 elaborates on the pattern. The pitch evolution progresses similarly to rhythm.

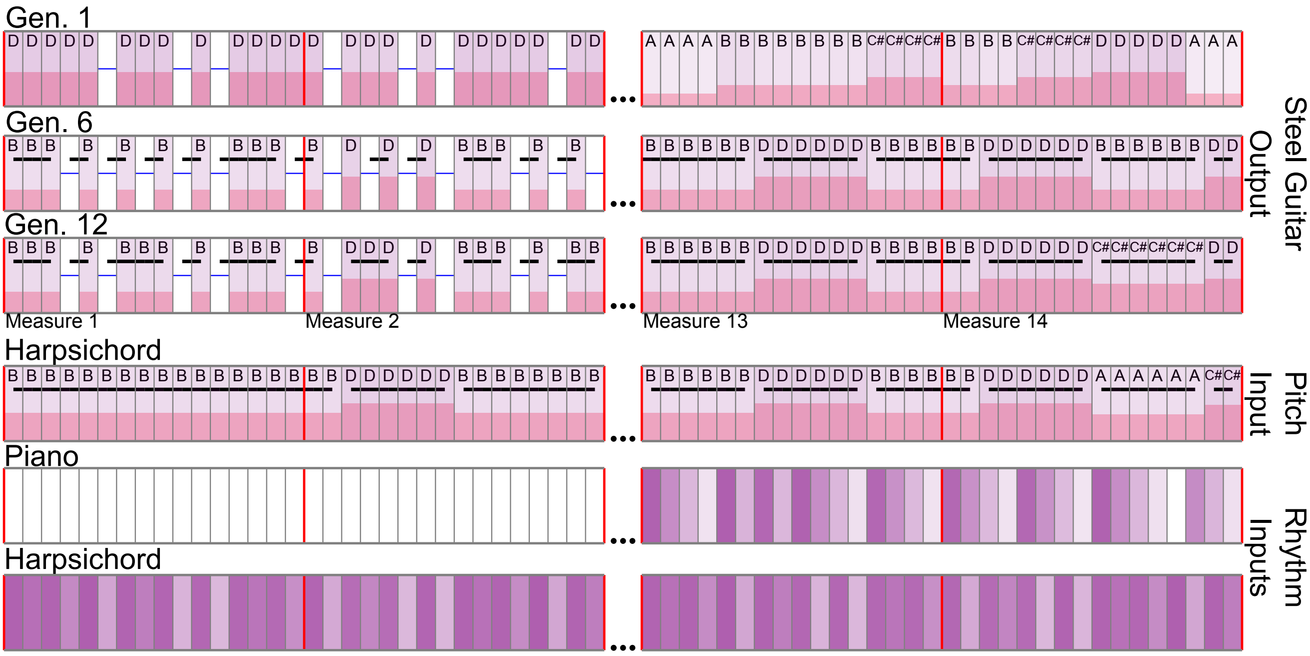


Figure 4: Evolutionary Accompaniment Sequence for Bad Girl's Lament. Evolved steel guitar accompaniment for generations 1, 6, and 12 of Bad Girl's Lament is shown at top, followed by the pitch and rhythm inputs to the CPPN from the scaffold. Each accompaniment can be heard at `http://eplex.cs.ucf.edu/fsmc/gecco2011/`. The type of instrument played in the scaffold is noted at left (e.g. a harpsichord is one channel in the rhythm scaffold. The increase in harmonic and rhythmic sophistication between generations 1 and 12 is apparent in the progression at top. In addition, the relationship (e.g. in note transition points) between the scaffold and accompaniment can also be observed.

From generation 1 to 6 many pitches change, but generations 6 and 12 differ in pitch by only a few choice notes.

While the generated pitch sequences in Bad Girl's Lament improve over twelve generations (as the listener study will confirm), some scaffolds are less amenable to guiding the system towards acceptable accompaniment, which can lead to fatigue. For example, figure 5 shows accompaniment for the folk song Scarborough Fair. The inputs for this selection are the oboe, guitar, clavi, and cello, and the output is a trombone. However, unlike with Bad Girl's Lament, wherein dissonance is less common even in the first generation, this first-generation accompaniment plays dissonant notes, like the C# in measure 9 on ticks 5 through 16. While dissonance is not inherently bad, without resolution, the pitches sound poorly chosen. For instance, the accompaniment would sound better by playing a G note instead of the C# note, thus resolving tension. While FSMC shows how much can be accomplished through functional scaffolding with little musical knowledge, this problem suggests a potential future opportunity to research tighter constraints on pitches while still preserving the functional relationship at the heart of FSMC.

5.2 Listener Study Results

The results from the 60-person listener study, which focused on the same IEC-evolved accompaniments for Bad Girl's Lament from the previous section, are shown in table 1. The first entry is the control example, which sounds intentionally artificial. As expected, it is rated significantly worse than every other example in the survey (at least $p < 0.05$ for all pair-wise comparisons with Student's t-test). This result establishes that listeners likely understood the questions in the survey.

MIDI Name	**Mean**	**Std. Dev.**
Poor Control	4.35	1.93
BGL without Accompaniment	7.30	1.85
BGL, Generation 1	5.15	2.20
BGL, Generation 6	6.07	1.96
BGL, Generation 12	6.83	1.98

Table 1: Perceived Quality by Survey Participants. This table shows the average ratings and the mean and standard deviation for the control and four Bad Girl's Lament (BGL) MIDIs. The MIDI names are on the left while the average ratings are on the right.

Importantly, generation 6 is judged significantly higher quality than generation 1 ($p < 0.05$) and generation 12 is judged significantly better than generation 6 ($p < 0.05$). Thus the participants felt that the progression from generation 1 to 12 indeed exhibits continual improvement in quality.

Furthermore, although the original MIDI without accompaniment is judged significantly better than generation 6 ($p < 0.001$), it is *not* judged significantly better than generation 12. Thus evolution guided by the human user eventually achieves in a short number of generations a level of quality indistinguishable from the quality of the original.

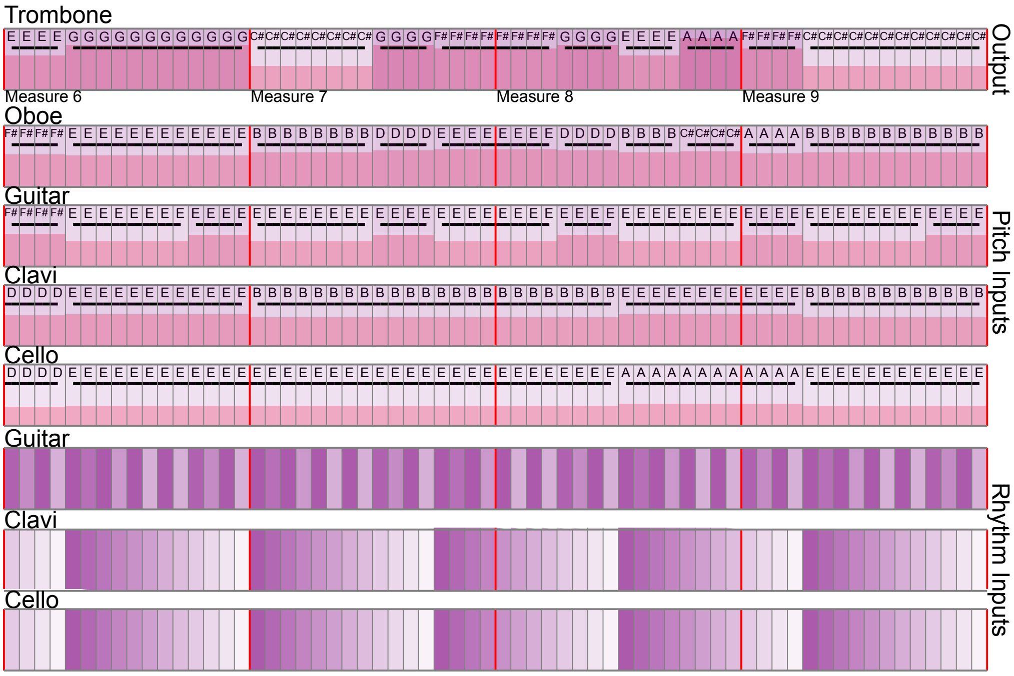


Figure 5: Dissonant Accompaniment for Scarborough Fair. This image shows an accompaniment from the first generation evolved with the Scarborough Fair scaffold. The accompaniment can be heard at `http://eplex.cs.ucf.edu/fsmc/gecco2011/`. Unlike with Bad Girl's Lament, for which many accompaniments sound harmonious in the first generation, Scarborough Fair produces some that are dissonant like this one. Thus one potential improvement for FSMC is more intelligent post-processing of the pitch outputs.

6. DISCUSSION AND FUTURE WORK

From the results of the listener study, it is apparent that IEC was effective in this case for navigating the search space induced by FSMC. The average ratings from generation 1 and generation 12 significantly improved while the quality of the original piece is indistinguishable from that in generation 12, demonstrating that IEC yielded a significant subjective improvement that ultimately re-approached the quality of the original song, yet now with a new accompaniment added.

However, although FSMC created an accompaniment of indistinguishable quality from a human composition, FSMC is not ultimately intended for accompaniment generation from complete songs that are already good. The results in this paper were accompaniments generated from the already-full compositions Bad Girl's Lament and Scarborough Fair because they meet a threshold of quality recognized in folk music. Thus accompaniment quality could be assessed relative to the original benchmark in different stages of development. Yet once this capability is established (through this study and others), eventually the aim will be to show that the quality of *incomplete* songs is actually *improved* by adding evolved accompaniment. Of course, it was difficult to improve the compositions in this study because they were already complete and well-regarded.

Because the listener study shows that it can take as few as 12 generations to reach an area of the search space of viable accompaniments, it is plausible to infer that the FSMC method is not forcing the user to search the entire space of possible accompaniments, which would be onerous. Both the CPPN representation and the scaffold help to capture the human essence of a pre-existing song, which the accompaniment can transform to sound plausible. In effect, FSMC "steals" the quality inherent in the scaffold and then manipulates it to create something new, thereby feeding off the skill of the human originator. Therefore, the FSMC method potentially can help users to find appealing accompaniments faster than could be found without such a method. This principle should even work with more complex musical pieces because the complexity in the scaffold will still be reflected in the output of the CPPN that transforms it.

Future work will focus on preserving the functional relationship between scaffold and accompaniment while at the same time imposing further constraints or post-processing on the pitch output, which is sometimes dissonant (as in the Scarborough Fair example). While dissonance is a problem, the *relative* movement and rhythm of pitch patterns output by CPPNs for Scarborough Fair are still compelling, suggesting that dissonance can potentially be treated as an orthogonal problem that is solved by manipulating but not removing the framework established in this paper.

7. CONCLUSION

This paper introduced functional scaffolding for musical composition (FSMC), a method for generating accompaniments from existing compositions, or *scaffolds*. Represented by CPPNs, patterns among pitches and rhythms in the scaffold are transformed yet respected. Elaborating on these patterns through IEC gives inexperienced composers the opportunity to explore accompaniment space for their own compositions. While future work will naturally focus on improving the method further, FSMC in effect opens up a new direction in research on evolutionary music generation by providing a succinct and effective theory based on a simple principle from which to build, i.e. that the different parts of a song are functionally related.

Acknowledgments

This work was supported in part by the National Science Foundation under grant no. IIS-1002507. Special thanks to Barry Taylor for granting special permission to utilize his own MIDI productions of folk music in this work. Barry Taylor originally sequenced Bad Girl's Lament and Scarborough Fair (without the new accompaniment).

8. REFERENCES

[1] Gérard Assayag, Georges Bloch, Marc Chemillier, Arshia Cont, and Shlomo Dubnov. Omax brothers: A dynamic topology of agents for improvization learning. In *Proc. of the 1st ACM workshop on Audio and Music Computing Multimedia*, pages 125–132. ACM, 2006.

[2] Kjell Bäckman and Palle Dahlstedt. A generative representation for the evolution of jazz solos. In *Proceedings of the Sixth European Workshop on Evolutionary and Biologically Inspired Music, Sound, Art and Design (EvoMUSART 2008)*, 2008.

[3] John A. Biles. GenJam in perspective: A tentative taxonomy for GA music and art systems. *Leonardo*, 36(1):43–45, 2003.

[4] Nick Collins. Interactive Evolution of Breakbeat Cut Sequences. In *Proceedings of Cybersonica, Institute of Contemporary Arts*, London, England, 2002.

[5] R. Dawkins. *The Blind Watchmaker.* Longman, Essex, U.K., 1986.

[6] D. A. Hart. Toward greater artistic control for interactive evolution of images and animation. In *Proceedings of the Fifth European Workshop on Evolutionary and Biologically Inspired Music, Sound, Art and Design (EvoMUSART 2007)*, volume 4448, pages 527–536, Valencia, Spain, 2007. Springer-Verlag.

[7] Amy K. Hoover, Michael P. Rosario, and Kenneth O. Stanley. Scaffolding for interactively evolving novel drum tracks for existing songs. In *Proceedings of the Sixth European Workshop on Evolutionary and Biologically Inspired Music, Sound, Art and Design (EvoMUSART 2008)*, 2008.

[8] Amy K. Hoover and Kenneth O. Stanley. Exploiting functional relationships in musical composition. *Connection Science Special Issue on Music, Brain, & Cognition*, 21(2):227–251, June 2009.

[9] B. Johanson and R. Poli. GP-music: An interactive genetic programming system for music generation with automated fitness raters. *Proceedings of the Third Annual Conference: Genetic Programming*, pages 181–186, 1998.

[10] Kris M. Kitani and Hideki Koike. Improvgenerator: Online grammatical induction for on-the-fly improvisation accompaniment. In *Proceedings of the 2010 Conference on New Interfaces for Musical Expression (NIME 2010)*, 2010.

[11] P. Machado and A. Cardoso. All the truth about NEvAr. *Applied Intelligence Special Issue on Creative Systems*, 16(2):101–118, 2002.

[12] Artemis Moroni, Manzolli Jônatas Manzolli, Fernando Von Zuben, and Ricardo Gudwin. Vox populi: An interactive evolutionary system for algorithmic music composition. *Leonardo Music Journal*, pages 49–54, 2000.

[13] G. L. Nelson. Sonomorphs: An application of genetic algorithms to growth and development of musical organisms. In *4th Biennial Art and Technology Symp.*, pages 155–169, March 1993.

[14] H. Nishino, H. Takagi, S. Cho, and K. Utsumiya. A 3d modeling system for creative design. In *Proceedings of the 15th International Conference on Information Networking*, pages 479–487, 2001.

[15] J. Romero and P. Machado, editors. *The Art of Artificial Evolution: A Handbook on Evolutionary Art and Music.* Springer, 2007.

[16] Ribamar Santarosa, Artemis Moroni, and Jônatas Manzolli. Layered genetical algorithms evolving into musical accompaniment generation. In *Proceedings of the Fourth European Workshop on Evolutionary and Biologically Inspired Music, Sound, Art and Design (EvoMUSART 2006)*, pages 722–726, 2006.

[17] Jimmy Secretan, Nicholas Beato, David B. D'Ambrosio, Adelein Rodriguez, Adam Campbell, Jeremiah T. Folsom-Kovarik, and Kenneth O. Stanley. Picbreeder: A case study in collaborative evolutionary exploration of design space. *Evolutionary Computation*, 2011.

[18] Jimmy Secretan, Nicholas Beato, David B. D'Ambrosio, Adelein Rodriguez, Adam Campbell, and Kenneth O. Stanley. Picbreeder: Evolving pictures collaboratively online. In *CHI '08: Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pages 1759–1768, New York, NY, USA, 2008. ACM.

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

[20] Kenneth O. Stanley and Risto Miikkulainen. Evolving neural networks through augmenting topologies. *Evolutionary Computation*, 10:99–127, 2002.

[21] Kenneth O. Stanley and Risto Miikkulainen. Competitive coevolution through evolutionary complexification. *Journal of Artificial Intelligence Research*, 21:63–100, 2004.

[22] H. Takagi. Interactive evolutionary computation: Fusion of the capabilities of EC optimization and human evaluation. *Proceedings of the IEEE*, 89(9):1275–1296, Sep 2001.

[23] Belinda Thom. Learning models for interactive melodic improvisation. In *Proceedings of the International Computer Music Conference*, pages 190–193, 1999.

[24] N. Tokui and H. Iba. Music composition with interactive evolutionary computation. In *Proceedings of 3rd International Conference on Generative Art*, Milan, Italy, 2000.

[25] T. Unemi. SBART 2.4: breeding 2D CG images and movies and creating a type of collage. In *Knowledge-Based Intelligent Information Engineering Systems, 1999. Third International Conference*, pages 288–291, Adelaide, SA, Dec 1999.

[26] L. World. Aesthetic Selection: The Evolutionary Art of Steven Rooke. *IEEE Computer Graphics and Applications*, 16(1):4–5, 1996.