

Computational Creativity in Popular Music Composition

by

Paul M. Bodily

A dissertation proposal submitted to the faculty of

Brigham Young University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Computer Science

Brigham Young University

December 2015

ABSTRACT

Computational Creativity in Popular Music Composition

Paul M. Bodily

Department of Computer Science, BYU
Doctor of Philosophy

The field of computational creativity (CC) is concerned with “the philosophy, science and engineering of computational systems which, by taking on particular responsibilities, exhibit behaviours that unbiased observers would deem to be creative” [6]. In the domain of music, much has been attempted to endow computers with the ability to express creativity. The vast majority of this work has been in the classical and jazz genres, with only a few attempts being made at algorithmic composition of popular music. Popular music represents a number of unique challenges including the addition of lyrics and a global structure of related verse-chorus segments which are essential components in this genre.

We propose to investigate the problem of algorithmic composition in popular music in the form of generating novel pop leadsheets (i.e., written chords, melody, lyrics). We do so using the paradigm of a graphical model which models independence assumptions between the various subtasks of composition in order to focus on each subtask individually. We intend to use a combination of data-driven and rule-based approaches. Inasmuch as creativity in humans is generally defined together with an inspiring idea or mood, we likewise incorporate such a module in our system. We will demonstrate creativity in our artifacts via demonstration of novelty, surprise, and value using a combination of fitness functions and human surveys.

1 Introduction

There is a great deal of interest in endowing computers with some form of artificial intelligence (AI), that is, the ability to perform “tasks, which, if performed by a human, would be deemed to require intelligence” [34]. Examples include the ability to recognize images, to interpret natural language, and to navigate a vehicle, to name a few. Effective AI systems enable individuals and societies to automate and delegate work to solve specific, well-formed problems.

The realm of human intelligence, however, encompasses a broader range of problem-solving that includes areas of endeavor where “notions of optimality are *not* defined” [11] (emphasis added). We refer to this as creativity. Consider, for example the creativity exhibited in cooking. A good chef employs a great deal of intelligence to concoct a delicious dessert, but unlike a winning outcome in a game of chess, there is no clearly defined notion of an optimal dessert.

We must not confuse an absence of clearly defined goals with a low estimation of the intelligence required to accomplish the task. We know a good dessert when we taste one and we value it. Societies pay top dollar for beautiful art, inspiring music, and illuminating works of literature. “Creative people and their contributions to cultural progression are highly valued” [6].

Creative endeavor may, in fact, require *more* than intelligence insofar as it necessitates not only the ability to generate solutions but also the capacity to be self-critical. As has been noted, “a poet with no critical ability to judge its own work... is no poet at all” [6].

And yet, despite its requisite self-awareness, creativity itself remains a mystery. Neither those who exhibit creativity nor those who study the psychology of creativity know how original ideas come about [2, 14]. Some advocate that creativity is purely a matter of generating creative artifacts. Others say that the process by which the artifacts arise is of primary concern. There is no clear line demarcating that which is creative from that which is not.

Computers can help us to understand ourselves and our creativity [2]. In the strong sense, the field of computational creativity (CC) is dedicated to elucidating the creative ability of humans and endowing it upon computers. It’s not simply using computers to be creative ourselves; it’s making computers to be creative themselves. As such, CC has been referred to as ‘creativity at the meta-level’ [3].

We are familiar with the measurable, tangible benefits of endowing a computer with intelligence. What benefit is there to endowing a computer with creativity, meaning the ability to perform “tasks, which if performed by a human, would be deemed *creative*” [34] (emphasis added)? Though some bias exists against machines being capable of creativity, researchers remain hopeful that the day will come when CC artifacts will be valued in their own right. This change will be more quickly realized as the quality of artifacts created by CC systems increases. Creativity in the form of game and game-level generation, and math theorem discovery has already gained acclaimed success [6]. But imagine a day when successful military tactics, winning sports strategies, exquisite cuisine, prize-winning art, and inspiring literature are all artifacts of CC systems.

Beside the artifacts, however, it is the “the nature of creative mental processes” [14] that interests us. “All forms of expertise are a proper concern for anyone interested in how the mind works” [14]. In some sense, creativity is a crowning characteristic of humanity, not only for its ability to aid in survival, but also for its role in the development of culture and communication.

Music and natural language are two characteristically human attributes. Of particular interest is the creative ability to write lyrical songs, a creative endeavor requiring both music *and* natural language. Though each has an extensive history in AI (and CC), strangely their overlap is a relatively unexplored domain. Lyrical song-writing has the capacity to generate unique value and inspiration that neither the lyrics nor the music are capable of expressing individually [18]. As [33] explains, “independent creation of an excellent piece of music and a great text does not necessarily

result in a good song.” When properly synchronized, the affect of their combination is more than the sum of their parts. The universality of this mode of expression is a testament to its power to captivate, inspire, delight, and educate the human mind.

Furthermore, lyrical song-writing poses its own set of unique challenges. Linguistic challenges include: developing a cohesive lyrical text under the organizational constraints of (occasionally complex) rhyme schemes; text composition within verse/chorus structures that is both independently-meaningful and globally-coherent; and use of effective metaphors, dramatic tag lines, and other common lyrical elements. Musical challenges include: generation of melody to appropriately match and emphasize lyrical syllables; shaping of melody and harmonization to effect the mood expressed in the lyrics; the creation of singable melodies (both in terms of range and structure); and, as with lyrics, the synthesis of melody and harmony that is independently-meaningful and globally-coherent.

To be clear, our interest is in the generation of written, lyrical songs, with less regard to how these songs are acoustically performed. Our focus will be on pop music, insofar as pop music represents the genre to which the majority of today’s headphones and earbuds are tuned.

Pop music is a genre with unique opportunities for CC research. Popular music is predominantly defined by lyrical compositions. Unlike generating lieds [33] or jazz leadsheets [25], pop songs generally have a unique hierarchical structure: multiple, connected verses that develop a common theme, interspersed with choruses or other segments. A wealth of data is available for pop music research in the form of online databases containing song-specific lyrics, harmony, structural organization, and metadata leveraged from crowd-sourcing efforts. Pop music is an emotionally charged genre—a characteristic which is heightened by the addition of lyrics—which provides an excellent medium through which to examine the generation of artifacts with *musical meaning* (i.e., referencing an inspiring theme, mood, or intention). This referential or descriptive aspect of music is a ubiquitous element of human creativity [26].

A leadsheet for jazz music has been previously defined in [25]. However, this previous definition requires adaptation in order to include lyrics. We therefore define a lyrical leadsheet as two “parallel” sequences: one that contains chord labels; a second that contains notes; and a third that contains lyrics. A *chord label* is a triple $\langle R, T, d \rangle$, where R is a pitch-class (e.g., C), T is a chord type (e.g., ‘M7’), and d is the duration. A *note* is also a triple $\langle p, d, s \rangle$, where p is the pitch (e.g., a MIDI-pitch between 0 and 127 or a rest), d is the note duration, and s is the lyrical syllable associated with the note. A lyricized leadsheet is a pair $\langle C, N \rangle$ where $C = (C_1, \dots, C_k)$ is a temporal sequence of chord labels and $N = (N_1, \dots, N_l)$ is a temporal sequence of notes. C and N have the same total duration. We assume a one-to-one relationship between non-rest notes and syllables.

Our goal is to develop a system that addresses both musical and linguistic challenges to generate lyrical leadsheets, based on an inspiring theme or idea. In the interest of time, we expect to devote a greater portion of our energy to the musical challenges without wholly disregarding the lyrical challenges.

2 Related Work

Much of the previous work done in AI with relation to music and language composition comes to bear on solutions to the lyrical song-writing task. Following a brief review of previous work in these more general subfields of AI, we look at what others have attempted with respect to the problem of lyrical song-writing. The remainder of the document will then detail the implementation (and validation) of a novel computationally-creative lyrical song-writing system.

2.1 Research Area Overview

Since the dawn of the computer age, AI researchers have been interested in the problem of algorithmic composition in various forms. The breadth of problems that have been addressed ranges from composing melodies, to harmonizing existing melodies, to live jazz improvisation. Despite this breadth, the range of solutions is somewhat narrower. These approaches represent a valuable toolset in considering the generation of lyrical pop song generation. They include knowledge-based systems, grammar-based systems, Markovian systems, evolutionary systems, and machine learning systems ([26] provides a more complete survey).

David Cope in the 80's and 90's developed a system entitled Experiments in Musical Intelligence (EMI) with the goal of generating new compositions in the styles of various composers [8]. EMI's widely-acclaimed success is rooted in the extraction and recombination of patterns in the works of a selected composer. This approach is representative of knowledge-based systems, which employ explicit reasoning via sets of rules or constraints. The learning of stylistic patterns might also be considered a generative grammar approach to composition, where the grammar is induced from an artist's collective works. Cope's EMI was largely constructed to emulate classical and ragtime composers, where the domain is devoid of lyrical melodies.

Mark Steedman is known for using manually crafted grammars for generating twelve-bar jazz chord sequences. His 1984 paper demonstrates how the variety of chord progressions manifest in twelve-bar jazz tunes can all be explained using a set of hand-crafted rules [31]. Johnson-Laird also employ a grammar for the generation of jazz chord sequences [14].

Markov models represent another common approach to algorithmic composition. Ames demonstrates the computation of transition matrices of varying orders in order to predict what comes next in a sequence of states given the preceding context [1]. States may be notes, chords, or even phrase expressions.

As an augmentation to the Markov approach, Conklin and Witten describe multiple view-point systems, which allow several facets (e.g., duration, pitch, volume) of a musical expression to be encoded as a single token, rather than modeling each aspect of the music with a different model [7].

CHORALE is another commonly cited knowledge-based system [10]. Designed to harmonize Bach chorales, this system employs multiple viewpoints in its layout of 270 (fairly complex) rules. As is common with rule-based systems, an exhaustive set of rules is generally unachievable. Numerous exceptions remain unaccommodated.

Ramalho and Ganascia aim at simulating creativity via intention-based real-time jazz improvisation [28]. They leverage Pachet's notion of potential actions (PACTs) to describe intention in music [24]. A PACT represents a procedural (e.g., play a particular rhythm, play an ascending arpeggio) or property-setting (e.g., play loud, use the major scale) characteristic of music which serves to guide the composition according to specific intentions. This intention-based approach is supported by previous research which experimentally demonstrates the important role of contour in how tonal music is perceived [9].

2.2 Research-Related Work

The problem of generating lyrical songs is relatively new and unexplored, particularly in the pop music genre. We examine a few of the closest related projects here.

Pachet and Roy introduce the problem of generating musical leadsheets in the style of a particular composer [25]. They formally define a leadsheet as a parallel sequence of chords and notes. In the jazz domain, this would be considered a full composition. Musical generation is

accomplished with composer-specific training of Markov models. Their results indicate that lead-sheets generated in the style of a specific composer are correctly classified with accuracy comparable to classification of original leadsheets heldout from the training dataset. Though sufficient for jazz compositions, the additional complexity of pop music lyrics and verse/chorus structure generation remains unaddressed by this approach.

Toivanen et al. describe M.U. Sicus, a system which generates novel lied songs (art songs) in a particular mood with lyrics composed in the Finnish language [33]. Building on previous research into the generation of poetry, M.U. Sicus generates (in order) lyrics, rhythm, chords, and then melodies using a combination of grammar-induction and Markovian approaches. We plan to approach the problem in a similar manner, but with the added complexity of integrating verse/chorus structure. We plan to more carefully evaluate the alignment of strong versus weak syllables and the melody, which M.U. Sicus does not consider.

Monteith et al. generate melodic accompaniment for existing lyrics in order to generate new lyrical compositions [20]. They also create a data-driven dictionary from which rhythms are selected and melodies are generated according to a trained n -gram model. They generate melodies in three different genres for lyrics sampled from the same three genres and find that in many cases, the computer-generated melodies are found to fit better and be more pleasing than the original melodies as evaluated by a panel of human judges. The tasks of generating lyrics and generating harmony are not addressed.

Scirea et al. recently published the Scientific Music Generator (SMUG), which generates lyrics (inspired from academic papers) and accompanying melodies [30]. They sample lyrical sub-structures from a small database of existing pop songs which they use to construct new lyrical compositions. They use Markov models, modified via genetic algorithms, to generate sequences of notes and rhythms. Although they have a complete pipeline that generates lyrical compositions, the lyrical model lacks sufficient complexity to generate grammatical or semantic coherence.

Chuan and Chew generate and evaluate musical harmonizations that emulate style in popular music [5]. They generate the chords from the melody, using Markov chains to model the likelihood of chord patterns. Melody fragments that strongly imply certain chords are used as anchors which are connected using chord sequences generated by consulting the Markov chains. They focus on Western popular songs. They don't address lyrics, nor do they generate melodies.

3 Thesis Statement

A probabilistic graphical model which uses a combination of data-driven solutions and rule-based solutions for individual components of structure, chords, melody, and lyrics is able to generate novel lyrical leadsheets in the western popular music genre which demonstrate human-like creativity in terms of novelty, surprise, and value.

4 Project Description

Popular musicians are routinely questioned about their song-writing process. What comes first: The music? The lyrics? The chords? Answers vary sufficiently to suggest that there are numerous, equally successful approaches to the creation of new lyrical compositions. Though in some cases artists will insist that various elements are simultaneously generated, it is often the case that individual elements are composed under independence assumptions. For example, given a particular theme, an artist may entirely compose the lyrics before even attempting to compose the melody, or vice versa.

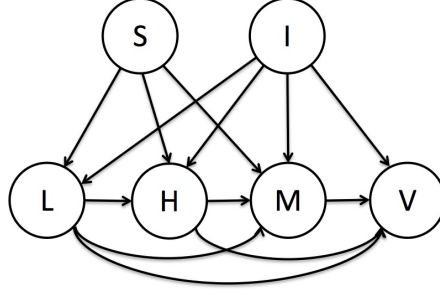


Figure 1: A probabilistic graphical model for pop music generation based on a set of independence assumptions. The model shown is comparable to a composer who strictly prefers to generate music in order of lyrics, chords, and then melody.

Figure 1 shows a graphical model that models a set of independence assumptions between the several subtasks of automated lyrical music generation in the genre of popular music. The set of independence assumptions will vary from one creative agent to another, and often a single creative agent may use a different creative model for different compositions. This particular set of independence assumptions represents the creative process that we intend to follow in this study. To model the lyrical leadsheet generation task in this way substantially simplifies the song-writing process insofar as solutions can be developed and evaluated modularly.

In plain English, the model essentially outlines how a particular song-writer would go about composing a novel pop song. First, a global song structure S (e.g., the sequence of verses, choruses, etc.) and a theme or inspiring emotion I are determined. Given these two elements, the system then generates (in order) the lyrics L , the harmony H , the melody M , and the voicing V . A lyrical leadsheet $\langle C, N \rangle$ is created by combining L and M to form N and letting $C = H$. M is generated under the constraint that it contains as many notes as L has syllables. S and I have only an indirect influence on the definition of the lyrical leadsheet. V refers to how the defined leadsheet is subsequently performed.

We plan to allow arbitrary selection of time signature between 3/4 and 4/4, as most popular music is written with these signatures. We will also assume that key signature remains constant for a given composition and is chosen as a function of the melody in order to make the song singable for a male tenor voice. In the following sections we define each of these elements in detail and discuss possible methods of implementation. In each case we propose a baseline (uncreative) solution and two creative solutions to each module. We initially plan to only implement one of the intelligent solutions. Should time allow (or if the first becomes for any reason infeasible), we will then consider the others. The dissertation will demonstrate that the combination of these solutions is sufficient to generating novel artifacts in the genre of popular music.

4.1 Structure

Repetition (e.g., of harmony, melody, or lyrics) is a critical component of successful pop music because it increases the ease with which information is processed by the brain [23]. We will refer to these patterns of repetition as the *structure* of the composition. Though not required to define a lyrical leadsheet, modeling structure—both at a global and substructural level—is an essential characteristic of successful song-writing.

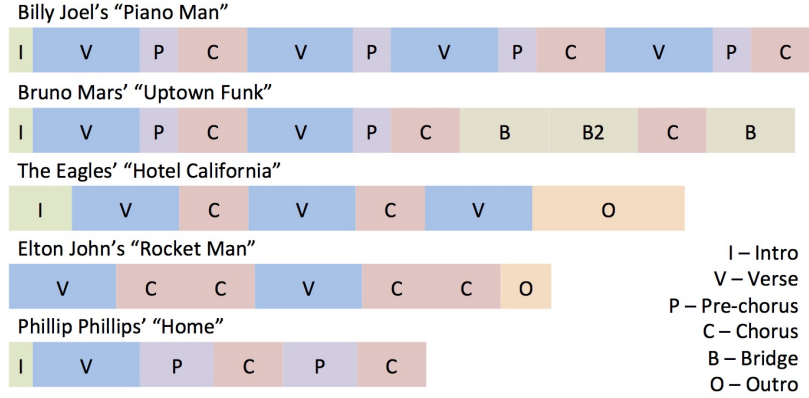


Figure 2: The inclusion and ordering of different song segments varies significantly, suggesting that global structure is a non-trivial creative aspect in the song-writing process.

4.1.1 Global Structure

The organization of a pop song into a sequence of segments¹ (e.g., intros, verses, pre-choruses, choruses, bridges, outros, etc.) is a unique element of pop music. To our knowledge this *global structure* (or *structure*), though studied extensively with respect to the segmentation of audio signal [13], has not been previously studied in the context of algorithmic composition. There is significant variety in the inclusion and ordering of the different segments (e.g., Figure 2) which adds to the novelty, surprise, and value of the composition. Consider, for example, the surprise/value of following a pre-chorus with something other than a chorus, as in Billy Joel’s “Piano Man”.

A baseline solution to generating global structure is to always use the same static structure (e.g., intro, verse, chorus, verse, chorus, bridge, chorus). This would be selected according to the structure that is most common in our training data.

One creative solution is to learn a distribution of possible structures from data. The idea of elucidating patterns from a corpus has been applied to generate rhythms [20] and lyrical substructures [30] but not at the level of global song structure.

A second creative solution is to learn a structure grammar from data where segments compose the set of terminals. This, too, has been tried with respect to other elements of composition (e.g., [31]) but not with global song structure. Compared to simply learning a distribution, a grammar has the added ability (for good or for bad) of creating structures that were not seen in the training set.

4.1.2 Substructure

There is really no limit to the degree of hierarchical repetition within a song (nor does it necessarily need to be strictly hierarchical). For example, consider Figure 3. “Piano Man” has four verses, each of which use the same harmonic progression, melody, and high-level rhyme scheme. Within this repeated verse segment, however, is layered additional repetition. The harmonic progression pictured in Figure 3 represents a substructure whose progression and rhyme scheme is repeated

¹As a working definition, a *segment* will refer to a unique (possibly repeated) subsequence of the composition. Thus a song with several verses will have a single “verse” segment. It should be noted that segment boundaries are not well-defined, and thus we might expect variability of segment labels by labeler. The exact labeling and assignment of segments is not essential to their effective use in recognizing a global structure of repetition.

	C		G/B		F/A		C/G
It's	a	pretty	good	crowd	for	a	Saturday
	F		C/E		D7		G
And	the	manager	gives	me	a	smile	
	C		G/B		F/A		C/G
'Cause	he	knows	that	it's	me	they've	been
	F		F/G		C		C
To	forget	about	life	for	a	while	

Figure 3: Several layers of structural repetition beyond the repetition of verse and chorus structure contribute to effective compositions. These lines from the first verse of “Piano Man” represent a substructure that is repeated twice in each verse and once in each chorus. Additional layers of substructure exist even *within* this substructure, demonstrating the effective use of hierarchical structure in pop music.

twice for each verse and once in each chorus. Furthermore the pictured progression itself repeats the sub-substructure of C—G/B—F/A—C/G (which is also present in the introduction). A third substructure example is present in the lyrics. Whereas each repetition of this substructure rhymes words at the ends of the 2nd and 4th lines (e.g., “smile” and “while”), five of the eight repetitions insert an additional substructural rhyme scheme in the 3rd line (e.g., “me” and “see”). Learning patterns of repetition at several layers (and not just at the highest or lowest level) enables a generative system to create novel combinations of layered repetition for novel compositions.

A baseline solution to generating substructure is to ignore it altogether. This would essentially result in songs where the only repeated elements are the segments themselves.

Any other solution will require the introduction of a substructural language capable of representing repetition of harmonic sequence and rhythm, melodic sequence and rhythm, lyrical phrases, and rhymed syllables. Though not all of these repeatable attributes are part of every repeated element, it is possible that some or all of them could be present together (e.g., in a chorus).

With the aid of the substructural language, we can begin to consider other ways of generating substructures. One creative solution is to generate abstract, parameterizable substructures for song segments, each of which is itself composed of other abstract, parameterizable substructures (cf. [16]). The parameters might include one or more abstract patterns of repetition, the repeatable attributes to which they apply, the number of times the pattern repeats within the substructure, and the depth of the hierarchy of abstract substructures within the given substructure.

A second creative solution is to learn a substructure distribution from data. To increase the power of our data, we would learn a distribution for each repeatable attribute in order to then generate new combinations of repeatable attributes for substructure elements.

4.2 Inspiring Idea

A critical element of creativity is the theme (e.g., loneliness, courage, despair), mood (e.g., joy, anger, sadness), or intent (e.g., to inspire, to intimidate) that motivates the generation of new artifacts. As shown in the model, this inspiring idea is assumed to be derived independent of solutions to the other song-writing subtasks. The basis for this assumption is that a song-writer is not likely to select his/her intent or mood based on the song that he/she is composing, but rather vice versa. The lyrics, harmony, melody, and voicing are all carefully crafted to reflect this motivating thought. Often, the value ascribed to a human-generated artifact is determined by the success of these secondary components in transmitting the inspiring idea.

For our purposes, we limit inspiring ideas to emotions, with each composition being limited

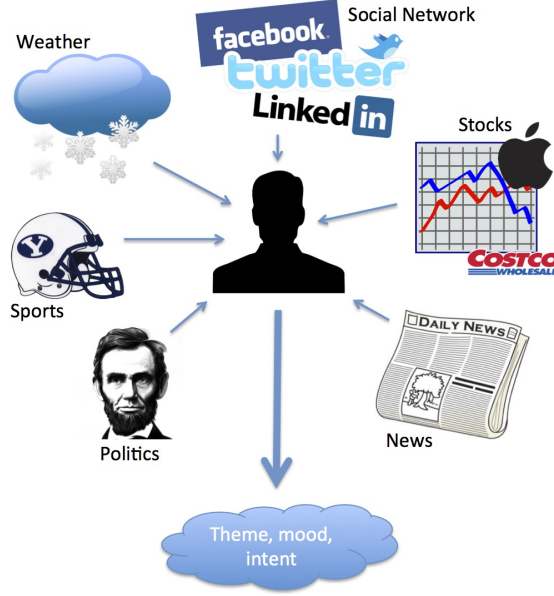


Figure 4: Creative artifacts are influenced by an inspiring idea that is affected by an agent’s environment, interests, and affiliations.

to a single emotion. At least two basic sets of emotions have been used in the CC literature [18, 19]. The first, outlined by Parrott [27], includes love, joy, surprise, anger, sadness, and fear. The second, outlined by Ekman [12], includes anger, disgust, fear, happiness, sadness, and surprise. We propose to use the latter insofar as distinguishing in pop music between love and emotions such as joy or sadness is challenging.

A baseline solution to generating an inspiring idea is to select randomly or according to some distribution (learned via sentiment analysis of other lyrical and musical inspirations) from the set of possible inspiring ideas.

One creative solution is to prompt the user for an inspiring idea from which to generate a lyrical composition. The system might then choose to either adopt the inspiring idea of the user, or alternatively to react to the user’s emotion (e.g., a song to cheer up a sad user).

A second creative solution is to select an inspiring idea based on interaction with an environment (e.g., via sentiment analysis of newspapers, sporting event outcomes, stock market, weather, social network posts, etc.). In this sense, our system would develop a *persona*, whose interests and affiliations imitate those of a human, change over time, and continually provide varied inspirations for its generative process (see Figure 4).

4.3 Lyrics

The generation of lyrics requires more than generating simply meaningful, grammatical text; lyrics must also be poetic (both structurally and linguistically) and must effectively communicate the inspiring idea. Regardless of the melody to be generated, the lyrics comprise the most fundamental component in estimating a song’s mood [4].

In composing a pop song (where the length is commonly between one hundred and three hundred words), generating semantically coherent lyrics is quite challenging, requiring consideration of lexical choice, syntax, and semantics. Numerous approaches to this problem have been and might

be proposed. Though well within the scope of our interest, time will likely limit how much we will be able to devote to this task. Thus, we propose here a few simple approaches which we hope will be reasonably effective, and anticipate a more thorough treatment of this task in post-doctoral research.

A baseline solution is to utilize lyrical templates derived from existing songs in which a certain subset of the lyrics are replaced (maintaining similar morphological form) to create a new set of lyrics (e.g., [32]). Though considered the baseline, this approach is likely to produce lyrics with greater semantic cohesion because it derives heavily from existing lyrical compositions.

One creative solution is to train an n -gram language model on lyrics from existing pop songs, thus capturing the nuanced grammatical and syntactical language of pop lyrics. We propose to conduct semantic analysis of these songs to predetermine the inspiring emotion of each song and train a different n -gram model for each of the six emotions. Data of this kind is easily accessible online. Numerous linguistic APIs exist to look up parts of speech², word roots, syllabic composition, and rhyming³ for a particular word. Other APIs exist to lookup groups of words matching specific criteria of length, spelling⁴, etc. Such a model will obviously be lacking semantic cohesion but represents a starting point.

A second creative solution is to train a generative model on an alternative text source. Other works have considered this approach, using academic papers [30] and newspaper articles [33]. Social media represents another rich resource for natural language inspiration. This approach might be augmented with the use of semantic webs for word substitution.

In the long run, we would be interested to investigate more complex NLP models that allow greater semantic cohesion between sentences and segments. There is also the issue of avoiding plagiarism that must be considered when generating novel text. Other ideas for future research include investigating means of maintaining a consistent point of view (e.g., first-person, etc.); incorporating metaphors (e.g., “candle in the wind”), similes (e.g., “cold as ice”), alliteration (“whisper words of wisdom”), personification (e.g., “my guitar gently weeps”), pop-culture references (e.g., “goodbye Norma Jean”), and possibly neologisms (e.g., “zip-a-dee-doo-dah”); and using words and phrases that are catchy (e.g., “you’re the inspiration”) or heavily charged (e.g., “just beat it”).

4.4 Harmony

Harmony, or the chord sequence of a composition, is dependent upon the inspiring idea, the global structure and the generated lyrics. Depending on the selected inspiring idea, the model governing chord sequence generation will vary. For example, Mihalcea and Strapparava found that key (which includes major or minor) is more capable than the notes themselves of predicting a song’s sentiment (as determined by a human listener) [18]. Clearly the pattern of repetition of chord sequences will depend on the structure and substructures of the global song segments. And the chord sequence lengths are a function of the syllable counts of the lyrical phrases.

We have chosen to generate harmony before melody based on the opinion that harmony is not only prerequisite to being able to evaluate a melody, but that a tonal melody, by definition, has its root in a particular harmonic context.

The baseline solution to creating harmony is to generate a single chord for the duration of the song (which has been known to occur in pop music). Though interesting songs might be generated from such a harmony, we would quickly look to more sophisticated solutions.

²e.g., <https://wordnet.princeton.edu/>

³e.g., <http://www.programmableweb.com/api/rhyming> and <http://rhymebrain.com/api.html>

⁴e.g., <http://www.datamuse.com/api/>

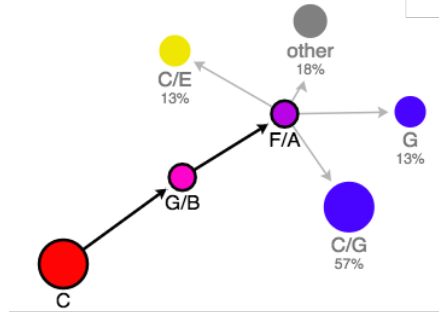


Figure 5: A graphical representation of a Markov model as learned and rendered by Hooktheory.com. In this representation, a context (of length 3) has been selected and the transition probabilities to the next chord are shown.

A more creative solution is to learn phrases of chord progressions from data, which might then be sampled at random to fill substructural elements.

A second creative solution is to use Markov models. Markov models, both single- and mixed-order, are commonly employed to generate musical sequence. Toivanen, for example, uses a second-order Markov chain trained on diatonic western classical music [33]. Similar models have been created for pop music and are available via online APIs⁵ (see Figure 5). The API provides start probabilities and transition probabilities, but notably absent are transitions to phrase- or song-endings. Another reason this model is insufficient is that it assumes a single model for all segments and for all inspiring ideas, which we will not assume.

For this reason, we plan to train our own diversified model. Tab data, labeled by segment and with the inspiring emotion, would serve very well for this purpose. Songs would first be transposed to a single key. Assuming a Markov order of m and n possible chord labels (including alterations and added intervals), n^m possible transitions exist. For even small values of m , this results in a large number of transitions (given that n is in the thousands). In reality, a very small subset of the possible transitions are actually observed. If after looking at the power of the model we observe there are still too many transitions, there are effective ways to reduce n . In the limit, we can ignore everything but the root note (e.g., C) and the chord quality (i.e., major or minor), thereby limiting n to 24. This practice is notably used in [31].

One concern that might be raised relating to our use of tablature data is that the quality of transcriptions is often suboptimal. This is most commonly due to lack of expertise and not lack of attention to detail, which results in *patterns* of errors in the data. We propose that such data is yet sufficient for our purposes as it still represents a human-generated creative artifact (though perhaps not that of the original composer). Consider, for example, that the tab used for the structure elucidation in Figure 7 contains a number of chords that we consider to have been incorrectly transcribed.

It is commonly noted that pop songs suffer from an over-repetition of certain chord progressions (e.g., I-IV-V). Although there are sound mathematical relationships which explain the stability of these chord sequences, it may be desirable to generate a wider variety of harmonies. One way to do this would be to equalize probabilities in the model according to some parameter, which in the limit would make all transitions (seen at least once in the data) of equal probability. Adjusting this parameter would then determine the amount of novelty and/or surprise in the

⁵<http://www.hooktheory.com/trends>

harmonic progression.

Repetitions of chords in phrases or groups of phrases will be pre-determined by substructural elements. In this way we hope to approximate what Keller 2014 calls “idiomatic chord progressions” which are chord sequences that repeat in places throughout a composition.

We will initially make the simplifying assumption of a single chord per measure and leave the consideration of multiple chords per measure (and their spacing) for future work.

4.5 Melody

Melody, like lyrics and harmony, is capable of (more or less) effectively conveying the inspiring idea. *Melody* consists of a sequence of notes, each of which has a pitch (or rest) and a duration. The range of valid pitches corresponds to the MIDI note numbers, though in practice we will limit the range further to achieve melodies whose range is transposable into a singable range (e.g., 43-67). The minimum duration (which we will consider) corresponds to a 1/48 note to allow for triplet sixteenth notes as well as sixteenth notes. The maximum duration will be a whole note. The task of generating rhythms is commonly separated as precursory to generating pitches. It may be that they can or should be treated jointly, but for the sake of simplicity, we intend to follow this common practice.

4.5.1 Rhythm

The sequence of note (and rest) durations is what is defined as *rhythm*. The particular rhythm of a melody determines in large measure which (corresponding) syllables receive emphasis. Rhythms that appropriately emphasize strong syllables contribute to a more natural pronunciation and comprehension of the lyrics. As an example, consider the line “I’M a LIT-tle TEA-POT” (traditionally sung in 4/4) in which each of the 4 beats of the measure coincide with a strong syllable in the lyrics (i.e., those capitalized). The pronunciation becomes more forced and the lyrics less comprehensible if, instead, we alter the rhythm to align the following capitalized syllables with beats 1-4: “i’m A lit-TLE tea-POT war-MER”. Good rhythms avoid placing weak syllables on the principle beats of the measure.

A baseline solution to the problem of generating rhythm is to randomly select note durations within a more conservative range of durations (e.g., 1/8 to 1/2 notes).

A more creative solution is to sample durations from a distribution of rhythms learned from data. We might also consider sampling a multivariate distribution which in addition considers the position in the measure for which the duration is being sampled.

Another creative solution is modeled by [20] who use a rhythm dictionary, learned from data, to generate new melodic rhythms. Their unique spin is to generate 100 rhythms in this way, use the CMU Pronunciation Dictionary to determine the stress patterns of the phonemes constituting the pre-generated lyrics, and then score each of the 100 rhythms based on how well the downbeats of each measure correspond with stressed syllables. In our estimation, this method might be improved upon by ensuring that, more than considering just the downbeat, as many stressed syllables as possible are aligned with beats. For this reason, we propose to use the stress patterns in the pre-generated lyrics to generate the rhythm, assigning durations according to a dynamic probabilistic model that gives highest probability to durations that would place the next strong syllable on the next beat.

4.5.2 Pitch

The generation of pitches to form interesting melodies has been the subject of significant research. A baseline solution to generating pitches simply chooses random pitches from those belonging to the triads of the accompanying chords.

A more creative solution to generating pitch is to use a Markov chain to probabilistically choose the next pitch based on the contextual sequence of preceding pitches. The model would also be conditioned on the position in the measure for which the pitch is being determined and the current chord in the harmonic progression.

A second creative solution, rather than than generating pitch note by note, is to first define a sequence of idiomatic expressions similar to Keller’s idiomatic chord progression bricks, Pachet’s PACTs, or Murray’s melody shapes [15, 21, 24]. Each of these expressions essentially defines a contour or feature that the group of notes at a particular point in the sequence should form. For example, a rising line, a descending arpeggio, an oscillation between two neighboring notes, etc. This can also be accomplished using mathematical regression.

The key signature of the composition will not be determined until after the melody has been generated. This is so that based on the melody, a key can be selected that maximizes singability within a determined vocal range.

Most effective compositions combine the melody and lyrics in a way that achieves an effect that is greater than the sum of parts. As an example, the contour or features in the melody can be made to accentuate the effect of the lyrics (this is called *prosody*). Consider such well-known lines as “*stop* in the name of love,” “somewhere over the rainbow way *up* high,” and “and I’m *free*, free *fallin’*.” In these lines, the italicized words are emphasized by the melodic features that they suggest. This has been minimally studied in the context of CC and is an issue we feel deserves attention.

Here, as with lyrics, we would do well to address the issue of avoiding plagiarism, meaning the verbatim copying of melodies from instances in the training data whether intentionally or inadvertently through the generative process. Though significant work might be invested in this vein (cf. [26]), it might be resolved by simply changing, adding, or deleting a few notes (either randomly or in accordance with a trained model) in order to create a novel melody.

4.6 Voicing

Having defined the lyrics, the harmony, and the melody, we have now generated a lyrical song in what is essentially defined as a *leadsheet*. Voicing refers to how the song is then performed, including instrumentation, dynamics, tempo, and how the various harmonic voices will (or will not) be represented. The decision of which notes in the harmony will be voiced and how they will complement the melody is a critical element that has as much of an impact on the perception of the song as the composition itself [26].

This problem has been addressed in various places in the CC literature but is not of primary interest to our study. For our purposes, the composition itself is the creative artifact and its interpretation by performers is itself another. Countless artists have covered the popular Beatles song “Yesterday,” each applying his/her own creative voicing of the original creative masterpiece.

In order to allow those unfamiliar with reading leadsheets to evaluate our artifacts, we plan to implement a simple voicing module consisting of a male tenor voice singing the lyricized melody to a simple piano accompaniment which voices the harmony. Evaluation will be strictly limited to the non-voiced elements.


```

A: C G Am C F C D G C G Am C F G C C F F
B: C G Am C F C D G C G Am C F G C C C G Am C F G C C F F C
C: G Am C F C D G C G Am C F G C C Am C D7 D7 Am C D7 D7 G G7 G G7
B: C G Am C F C D G C G Am C F G C C C G Am C F G C C F F C
D: C F F
B: C G Am C F C D G C G Am C F G C C C G Am C F G C C F F C
C: G Am C F C D G C G Am C F G C C Am C D7 D7 Am C D7 D7 G G7 G G7
B: C G Am C F C D G C G Am C F G C C C G Am C F G C C F F C
C: G Am C F C D G C G Am C F G C C Am C D7 D7 Am C D7 D7 G G7 G G7
B: C G Am C F C D G C G Am C F G C C C G Am C F G C C F F C
D: C F F
B: C G Am C F C D G C G Am C F G C C C G Am C F G C C F F C
C: G Am C F C D G C G Am C F G C C Am C D7 D7 Am C D7 D7 G G7 G G7
B: C G Am C F C D G C G Am C F G C C C G Am C F G C C F F C

```

Figure 7: Using a suffix tree like that in Figure 6, we algorithmically segmented the chord sequence of “Piano Man” based on repetitive subsequence. The blocks roughly represent the following segments: A—introduction; B—the first half of the verse or the entire chorus; C—the second half of the verse; D—the last of the chorus.

sequence were identified). The resulting segmentation of “Piano Man”, as seen in Figure 7, though unintuitive compared to a traditional verse-chorus structure, effectively captures a global structure of repetition (based solely on the chord sequence). This might suggest that global structure can be generated implicitly through hierarchical repetition of subsequences. We are interested in comparing this approach with the explicit modeling of global structure, suspecting that the latter will produce songs whose structure fits more easily into the range of expectation of most listeners.

Melody models cannot be trained from tablature data but will require an alternative data source, such as MIDI files. Several sites exist with freely available MIDI files for pop music.

5 Validation

Validation in CC research differs significantly from validation in AI. Whereas AI generally describes well-defined problems with optimal outcomes against which AI solutions can be ranked and/or adapted, typically no ideal outcome exists for endeavors in CC. CC artifacts and behaviors are validated by demonstrating some set of attributes. For example, Boden [2] defines creativity in terms of novelty, surprise, and value. Ritchie [29] evaluates creativity using novelty, typicality, and quality. Here we suggest ways in which we might demonstrate these attributes in the behavior of our song-writing system at the modular level.

As mentioned in the introduction, a creative system requires the ability to be self-critical. The creative process tends to require iterative development at each song-writing step. To this end, we propose to encode a fitness function for each individual module which is tuned to promote artifacts which best reflect the above-mentioned attributes of creativity. Each module will thus be designed to both *generate* and *evaluate* artifacts in order to select a single choice for the compositional element.

In words attributed to Igor Stravinsky, “lesser artists borrow, great artists steal.” *Novelty* is generally assessed as a real-valued function, expressing a degree of similarity to pre-existing artifacts. In considering global structure, novelty can be assessed according to how many times (if any) a generated structure (in its entirety) has been previously used (as per available data). In this case we are less concerned about computing similarity between non-identical structures because structure is highly reused or largely similar between compositions (and can at times vary only as

a matter of subjective opinion). With harmony, novelty can be determined as a product of the novelties of all chord subsequences. Using this function, novelty is computed for one or several fixed lengths (the length determining a specific granularity) and is also computed as an interpolation of several lengths (reflecting a more comprehensive novelty estimate). Novelty in lyrics and melody would be similarly evaluated.

Closely related to novelty is *typicality*, the notion that an object fits within the definition of a particular artifact domain. Though highly formulaic, pop music does not generally adhere to specific rules about what does and does not constitute a pop song. There is extensive variation in the lengths, ranges, and complexities of global structure, lyrics, harmony, and melody. Any constraints we might impose in this regard to favor typicality are more likely to represent statements of what is *valued* in a pop song (see below), not how it is defined. To this end, we will withhold from incorporating typicality measures in our modular fitness functions.

Surprise refers to the unlikeliness or the interestingness of an artifact or behavior [2]. A common approach to evaluating surprise is the use of information theory [17]. In essence, when generating a musical sequence (e.g., any of the modular elements of composition), the *expectation* can be computed for any of the possible successive symbols. Generations with high expectation are likely to be good artifacts but are in some sense less interesting. Controlled generation of less expected sequences is one way to introduce surprise to the listener. Surprise might thus be thought of as knowing the “rules” of the art and intelligently breaking them. We would like to have neither too much nor too little surprise. We will learn from data both the magnitude and patterns of surprise in successful popular music and use this range in our development of modular fitness functions.

Value, as Boden notes, is the characteristic in which “many arguments about creativity are rooted” [2] (emphasis added). Defining value (or *quality*, as [29] calls it) is more difficult than defining previous attributes, in part because it is so subjective; what one listener enjoys, another may hate and vice versa. Several rules have been suggested for producing successful songs (e.g., “get to the chorus within the first minute of the song”, “only use a pre-chorus if the verse is short”, “a bridge or instrumental solo should not come before the second chorus”), and as a first approach we will attempt to utilize these rules. But numerous, extremely successful counterexamples exist for each of these rules, encouraging the notion that perhaps we should explore breaking the rules.

Ultimately we are interested in what *people* think of the generated music. We plan to use human questionnaires to assess the perceived creativity in generated artifacts at each step of the modular development. For each module, we will generate several hundred artifacts and pick the most fit (as per the module’s fitness function) as input for subsequent modules. We will control the other variable components by assigning them values taken from existing (less well-known) pop music (e.g., as done in [20]). This way we can directly compare our generated artifacts for each component against A) other computationally creative systems which attempt to address the same tasks and B) other human-derived solutions. We are interested to see, given no knowledge about the composition’s composer, how the creativity perceived in our generated artifacts compares to that perceived in others. Questionnaires will involve questions (regarding cohesion, complexity, interestingness, novelty, etc.) about the overall composition as well as about the specific module being evaluated.

Often value is ascribed not only based on the artifact itself (cf. [29] for a counterargument) but also based on the process that brought it about [6]. Norton et al. [22] use a series of surveys in which the origin and creative process behind artifacts under comparison are A) masked, B) stated briefly, and C) given in detail. Their purpose is to consider the effects of responder bias given the different amounts of information. We propose conducting a similar series of surveys but with an added questionnaire which masks the origin (i.e., human vs computer) while still managing

to disclose information about the creative process. This would allow improved separation of the potentially confounding effects of bias towards computer systems and bias resulting from knowledge of the creative process.

Lastly, we are interested in how well we communicate through music. To this end we would propose implementing a questionnaire which seeks to determine for each of several generated songs which emotion is perceived. The responses can then be compared to the inspiring emotions used to originally generate the songs (cf. [19]). In a similar vein, we could tune fitness functions to create songs which demonstrate particular attributes (or lack thereof) and then use questionnaires to observe whether the applied filter achieves the desired perception. We are considering the possibility of staging a concert with live singers and a mix of human- and computer-generated music. Questionnaires would then take the form of an audience texting results during a live concert.

6 Thesis Schedule

- November 2016 - Get formatted tab data
- February 27, 2016 - Papers submitted to the International Conference on Computational Creativity (ICCC) 2016:
 - Summary of prospectus idea
 - “Structure elucidation and learning in western pop music”
 - “Computational creativity: to what end?”
- April 2016 - Paper on “Generating harmony in western pop music” submitted to the Journal of New Music Research
- September 2016 - Paper on “Generating melody in western pop music” submitted to the International Conference on Evolutionary and Biologically Inspired Music, Sound, Art and Design 2016 or Computer Music Journal
- January 2017 - Papers submitted to ICCC 2017:
 - “Generating lyrics in western pop music”
 - “Is all creativity computational?”
- January 2017 - Submit dissertation to committee
- February 2017 - Submit musical work to the 43rd International Computer Music Conference
- March 2017 - Dissertation defense
- April 2017 - Paper that puts it all together submitted to Digital Creativity, AI Magazine, or Minds and Machines

References

- [1] Charles Ames. The Markov process as a compositional model: a survey and tutorial. *Leonardo*, pages 175–187, 1989.
- [2] Margaret A Boden. *The Creative Mind: Myths and Mechanisms*. Psychology Press, 2004.
- [3] Bruce G Buchanan. Creativity at the metalevel: AAAI-2000 presidential address. *AI Magazine*, 22(3):13, 2001.

- [4] Chung Yi Chi, Ying Shian Wu, Wei Rong Chu, Daniel C Wu Jr, Richard Tzong Han Tsai, et al. The power of words: Enhancing music mood estimation with textual input of lyrics. In *Proceedings of the International Conference on Affective Computing and Intelligent Interaction and Workshops*, pages 1–6. IEEE, 2009.
- [5] Ching-Hua Chuan and Elaine Chew. Generating and evaluating musical harmonizations that emulate style. *Computer Music Journal*, 35(4):64–82, 2011.
- [6] Simon Colton and Geraint A Wiggins. Computational creativity: the final frontier? In *Proceedings of the European Conference on Artificial Intelligence*, volume 12, pages 21–26, 2012.
- [7] Darrell Conklin and Ian H Witten. Multiple viewpoint systems for music prediction. *Journal of New Music Research*, 24(1):51–73, 1995.
- [8] David Cope. *Experiments in Musical Intelligence*, volume 12. AR editions Madison, WI, 1996.
- [9] W Jay Dowling. Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, 85(4):341–354, 1978.
- [10] Kemal Ebcioglu. An expert system for harmonizing four-part chorales. In *Proceedings of the Association for the Advancement of Artificial Intelligence*, pages 784–788, 1986.
- [11] Arne Eigenfeldt, Adam Burnett, and Philippe Pasquier. Evaluating musical metacreation in a live performance context. In *Proceedings of the Third International Conference on Computational Creativity*, pages 140–144, 2012.
- [12] Paul Ekman. Facial expression and emotion. *American Psychologist*, 48(4):384–392, 1993.
- [13] Jonathan Foote. Automatic audio segmentation using a measure of audio novelty. In *Proceedings of the IEEE International Conference on Multimedia and Expo*, volume 1, pages 452–455. IEEE, 2000.
- [14] Philip N Johnson-Laird. Jazz improvisation: A theory at the computational level. In Peter Howell, Robert West, and Ian Cross, editors, *Representing Musical Structure*, pages 291–325. Academic Press, London, 1991.
- [15] Robert Keller, Alexandra Schofield, August Toman-Yih, Zachary Merritt, and John Elliott. Automating the explanation of jazz chord progressions using idiomatic analysis. *Computer Music Journal*, 37(4):54–69, 2013.
- [16] Dean M LeBaron, Logan A Mitchell, and Dan Ventura. Intelligent content generation via abstraction, evolution and reinforcement. In *Proceedings of the Artificial Intelligence and Interactive Digital Entertainment Conference*, pages 36–41, 2015.
- [17] Leonard B Meyer. *Emotion and Meaning in Music*. University of Chicago Press, 2008.

- [18] Rada Mihalcea and Carlo Strapparava. Lyrics, music, and emotions. In *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 590–599. Association for Computational Linguistics, 2012.
- [19] Kristine Monteith, Tony Martinez, and Dan Ventura. Automatic generation of music for inducing emotive response. In *Proceedings of the International Conference on Computational Creativity*, pages 140–149, 2010.
- [20] Kristine Monteith, Tony Martinez, and Dan Ventura. Automatic generation of melodic accompaniments for lyrics. In *Proceedings of the International Conference on Computational Creativity*, pages 87–94, 2012.
- [21] Skyler James Murray and Dan Ventura. Algorithmically flexible style composition through multi-objective fitness functions. In *Proceedings of the International Workshop on Musical Metacreation*, pages 55–62, 2012.
- [22] David Norton, Derrall Heath, and Dan Ventura. Accounting for bias in the evaluation of creative computational systems: An assessment of DARCI. In *Proceedings of the International Conference on Computational Creativity*, pages 31–38, 2015.
- [23] Joseph C Nunes, Andrea Ordanini, and Francesca Valsesia. The power of repetition: repetitive lyrics in a song increase processing fluency and drive market success. *Journal of Consumer Psychology*, 25(2):187–199, 2015.
- [24] François Pachet. Representing knowledge used by jazz musicians. In *Proceedings of the International Computer Music Conference*, pages 285–285, 1991.
- [25] François Pachet and Pierre Roy. Imitative leadsheet generation with user constraints. In *Frontiers in Artificial Intelligence and Applications*, volume 263, pages 1077–1078. IOS Press, 2014.
- [26] George Papadopoulos and Geraint Wiggins. AI methods for algorithmic composition: A survey, a critical view and future prospects. In *Proceedings of the AISB Symposium on Musical Creativity*, pages 110–117. Edinburgh, UK, 1999.
- [27] W Gerrod Parrott. *Emotions in Social Psychology: Essential Readings*. Psychology Press, 2001.
- [28] Geber Ramalho and Jean-Gabriel Ganascia. Simulating creativity in jazz performance. In *Proceedings of the Association for the Advancement of Artificial Intelligence*, volume 94, pages 108–113, 1994.
- [29] Graeme Ritchie. Some empirical criteria for attributing creativity to a computer program. *Minds and Machines*, 17(1):67–99, 2007.

- [30] Marco Scirea, Gabriella AB Barros, Noor Shaker, and Julian Togelius. SMUG: Scientific music generator. In *Proceedings of the International Conference on Computational Creativity*, pages 204–211, 2015.
- [31] Mark J Steedman. A generative grammar for jazz chord sequences. *Music Perception*, 2(1):52–77, 1984.
- [32] Jukka Toivanen, Hannu Toivonen, Alessandro Valitutti, Oskar Gross, et al. Corpus-based generation of content and form in poetry. In *Proceedings of the International Conference on Computational Creativity*, pages 175–179, 2012.
- [33] Jukka M Toivanen, Hannu Toivonen, Alessandro Valitutti, et al. Automatical composition of lyrical songs. In *Proceedings of the International Conference on Computational Creativity*, pages 87–91, 2013.
- [34] Geraint A Wiggins. A preliminary framework for description, analysis and comparison of creative systems. *Knowledge-Based Systems*, 19(7):449–458, 2006.

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a dissertation proposal submitted by

Paul M. Bodily

This dissertation proposal has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

Date

Dan Ventura, Chair

Date

Neil Thornock

Date

Mark Clement

Date

Christophe Giraud-Carrier

Date

Seth Holladay

Date

Quinn Snell
Graduate Coordinator