# Creative Leadsheet Generation for Popular Music

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

November 2015

ABSTRACT

Creative Leadsheet Generation for Popular Music

Paul M. Bodily
Department of Computer Science, BYU
Doctor of Philosophy

The field of computational creativity (CC) is concerned with the generation of artifacts that if exhibited by a human would be deemed creative. In the domain of music, much has been attempted to endow computers with the ability to express musical 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 of creativity 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 generation 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" [29]. Examples include the ability to recognize images, interpret natural language, and 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, 13]. 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]. The field of computational creativity (CC) is dedicated to elucidating the creative ability of humans and endow 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 from 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*" [29] (emphasis added)? Though significant 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" [13] that interests us. "All forms of expertise are a proper concern for anyone interested in how the mind works" [13]. 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 a long extensive history in AI (and CC), 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 [16]. As [28] explains, "independent creation of an excellent piece of music and a great text does not necessarily result in a good song." When properly synchronized, their affect of their combination is more than the sum of their parts. The universality of this 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; and the creation of singable melodies (both in terms of range and structure).

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 [28] or jazz leadsheets [22], pop songs generally have a hierarchical structure: multiple, connected verses that develop a common theme, interspersed with choruses or other segments. Because of its popularity, large online pop music databases are available with song-specific lyrics, harmony, structural organization, and metadata leveraged from crowd-sourcing efforts. "Most of human music is referential or descriptive" [23]. An essential aspect of CC is that not only is an artifact produced, but that it has musical meaning, referencing an inspiring theme, mood, or intention [23]. Pop music, perhaps more than any other genre, expresses the breadth of the human emotional spectrum, thus making it an idea genre for the development of computationally creative systems.

Our goal is to develop a system that addresses both musical and linguistic challenges to produce artifacts, given 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.

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 this work, we look at what others have attempted with respect to this specific problem. The remainder of the document will then detail the implementation (and validation) of a novel computationally-creative lyrical song-writing system.

## 2 Related Work

### 2.1 Research Area Overview

Since the beginning of computers, researchers were 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 ([23] 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 artists 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 also known for his use of 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 [27]. Johnson-Laird also employ a grammar for the generation of jazz chord sequences [13].

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 viewpoint systems, which allows 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 rule set is generally unachievable for which numerous exceptions must be accommodated.

Ramalho and Ganascia aim at simulating creativity via intention-based real-time jazz improvisation [25]. They leverage Pachet's notion of potential actions (PACTs) to describe intention in music [20]. 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 [21]. 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 leadsheets 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 lieds (art songs) in a particular mood with lyrics composed in the Finnish language [28]. Building off of 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 [18]. They also use a corpus-induced grammar 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 lyrics are not addressed.

Scirea et al. recently published the Scientific Music Generator (SMUG), which generates lyrics (inspired from academic papers) and accompanying melodies [26]. They sample lyrical substructures 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 generate and evaluate musical harmonizations that emulate style in pop songs [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 popular music leadsheets 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 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 music 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 song-writer would go about composing a novel pop song. First, a theme or inspiring emotion $I$ and a global song structure (e.g., the sequence of verses, choruses, etc.) $S$ are determined. Given these two elements, the system then generates (in order) the lyrics $L$, the harmony or chord sequence $C$, the melody $M$, and the voicing $V$. We plan to allow arbitrary selection of time signature between 3/4 and 4/4, as most pop songs are written with these signatures. We will also assume that key signature remains constant for a

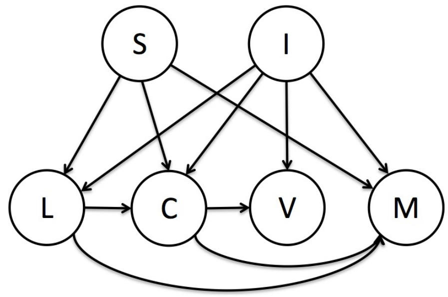


Figure 1: A probabilistic graphical model for pop music generation based on a set of independence assumptions. The model shown is comparative to a composer who strictly prefers to generate music in order of lyrics, chords, and then melody.

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 in detail and discuss possible methods of implementation. In each case we suggest what might be considered a simple solution to each module, and demonstrate that the combination of these simple solutions is sufficient to generating novel (albeit uninteresting) artifacts in the genre of pop music.

### 4.1 Inspiring Idea

A critical element of creativity is the theme, mood, or intent that inspires the generation of new artifacts. Examples include emotions such as joy, anger, or sadness. 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 creativity 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 to a single emotion. At least two basic sets of emotions have been used in the CC literature [16, 17]. The first, outlined by Parrott [24], 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 simple solution to generating an inspiring idea would be to select randomly or according to some distribution from the set of possible inspiring ideas. In addition to the simple solution, we propose two alternative solutions.

First, the system might 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).

Second, the system might 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 2).

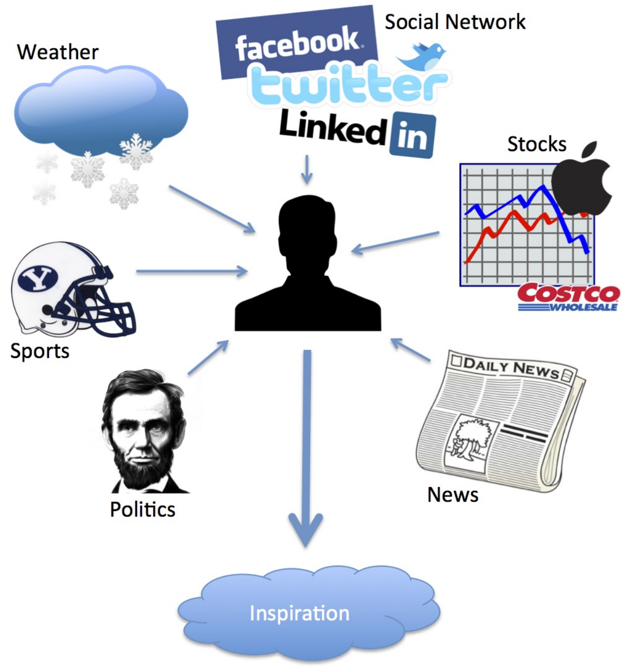


Figure 2: Creative artifacts are generated from an inspiring idea that is influenced by an agent's environment, interests, and affiliations.

### 4.2 Structure

The organization of a pop song into a sequence of segments[1] (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, has not been previously studied in the context of compositional generation. There is immense variety in the inclusion and ordering of the different segments (e.g., Figure 3) 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 simple solution to generating global structure would be to always use the same structure (e.g., intro, verse, chorus, verse, chorus, bridge, chorus). In addition to the simple solution, we propose two alternative solutions.

First, we might learn a distribution of possible structures from data. The idea of elucidating patterns from a corpus has been applied to generate rhythms [18] and lyrical substructures [26], but not at the level of global structure.

Second, a system might 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., [27]), but not with global structure. With respect to learning a simple distribution, a grammar has the added ability (for good or for bad) of creating structures that were not seen in the training set.

[1]For our purposes, a *segment* will refer to a unique element of the composition segment. Thus all verses represent a single verse segment. It should be noted that segments are not well-defined for which 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.

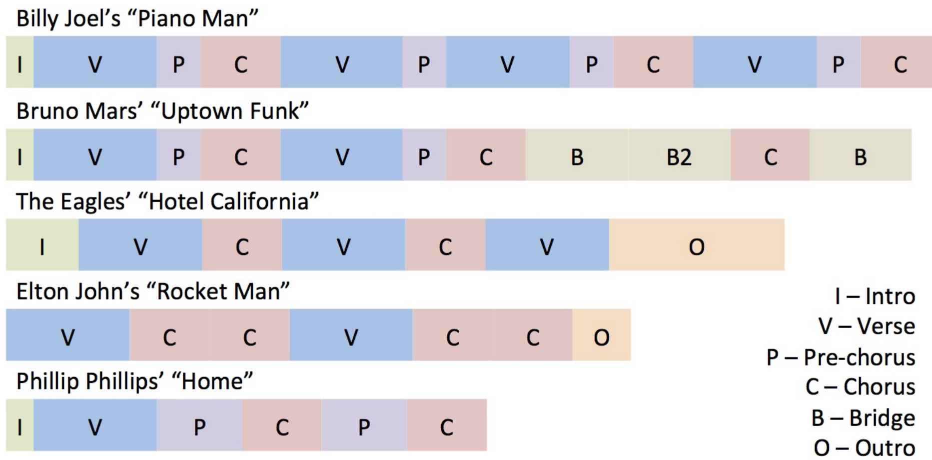


Figure 3: The inclusion and ordering of different song segments varies immensely, suggesting that global structure is non-trivial creative aspect in the song-writing process.

#### 4.2.1 Data

We currently have access to a dataset of over 800,000 pop tabs. These tabs, in free-form, unstructured text, generally contain lyrics and chord sequences. With the help of other resources, additional metadata might be added to include information such as the author, specific genre, and key.

Only a small percentage of these tabs have global structure that is clearly labeled, raising the question, "can global structure be inferred from just the chords and lyrics?" Labelling the sections of a song is second-nature to humans when listening to an acoustic recording. However, when only considering the chords and lyrics, the task is ill-defined. Unlike in many other genres, in pop music there are no rules which prescribe compositional or segment lengths. In some rare cases, pop songs may depart entirely from having any repeated structures whatsoever.

One approach to elucidating global structure from an unlabeled tab is to perform (fuzzy) pattern matching of chord sequences within the song to find repeated segments (e.g., verses). To this end, we have experimented using suffix tree (see Figure 4) that can be queried in constant time to find all instances of a chord subsequence of arbitrary length. Our preliminary results show promise, but we surmise that observing patterns in the lyrics (e.g., choruses) will improve results dramatically.

It may be unnecessary to elucidate global structure for the purposes of learning a distribution if, by some other means, the final artifact is found to contain this structure. Using a suffix tree like that pictured in Figure 4 but for all segments of "Piano Man", we incrementally identified the largest non-overlapping blocks of repeated subsequence until at least 90% of the sequence was contained in blocks (existing blocks were subdivided as new blocks with common sequence were identified). The resulting segmentation of Piano Man, as seen in Figure 5, though unintuitive as 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 to compare 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.

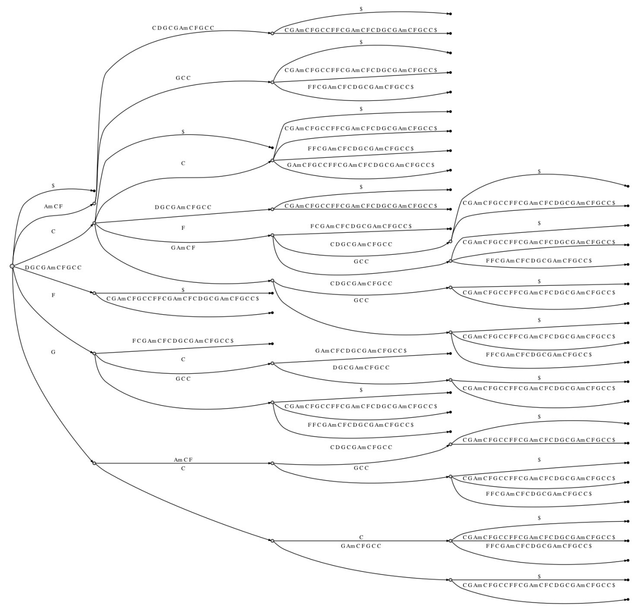

Figure 4: A graphical representation of the suffix tree created from the chord sequence of the first verse of Piano Man. Leaf nodes represent a unique position in the song where the subsequence (represented by the path from the root to the leaf node) begins. Such a suffix tree can be queried in constant time to find all instances of a chord subsequence of arbitrary length.

```
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 5: Using a suffix tree like that in Figure 4, 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.

Now looking back over the **years**,
And whatever else **appears**
I remember I **cried** when my father **died**
Never wishing to hide the **tears**

Figure 6: Several layers of structural repetition beyond the global segment structure contribute to effective compositions. Gilbert O' Sullivan's "Alone Again, Naturally" contains a hierarchical rhyming scheme which contributes to the uniqueness of the composition. Elements of the rhyme scheme are underlined and in bold.

#### 4.2.2 Substructure

The global structure alone is insufficient to produce songs of meaningful value. Many pop songs contain repetitive substructures, be it pairs or groups of notes in the melody, or pairs or groups of chords in the harmony. There are certainly substructures that place rhyming constraints on the lyrics. These substructural elements provide a level of repetition that has been shown to increase the success of pop music in terms of popularity because it increases the ease with which information is processed by the brain [19].

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 6. Although this verse may have an AABA form (where A and B represent rhyming lyrical phrases), further decomposition would reveal that B is itself composed of repeating elements (i.e., B = bb). This additional substructural element reveals additional repetition in the composition (e.g., AAbbA). It is also common for substructural elements (e.g., chord sequences) to be repeated *between* segments (see, for example, Figure 5).

The simplest approach 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 approach 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 is part of every repeated element, it is possible that some or all of them could be present together (e.g., in a chorus). The development of a substructural language is a high priority in the next stages of our research.

With the aid of the substructural language, we can begin to consider other ways of generating substructures. One approach may be to generate random substructures for song segments and then randomly define which repeatable attributes are actually repeated for each element in the substructure.

A second approach would be 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.

#### 4.2.3 Local Variation

Though pop music commonly exhibits a global structure and several substructures, the ability to vary structure at each repetition of a segment is a critical element of creativity that introduces novelty and surprise. Such localized variation can be generated through specific rule-based algorithms during later generative modules or as a post-generative step. We plan to consider this only after we have found meaningful solutions to the rest of the generative process.

### 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 generating a pop song (where the average length is commonly a few 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.

One approach we would like to try 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 would 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 lookup parts of speech[2], word roots, syllabic composition, and rhyming[3] for a particular word. Other APIs exist to lookup groups of words matching specific criteria of length, spelling, etc.[4]. Such a model will obviously be lacking semantic cohesion, but represents a starting point.

A second approach would be to train a generative model on an alternative text source. Other works have considered this approach, using academic papers [26] and newspaper articles [28]. 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., "moves like Jagger"), 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., "shut up and dance").

### 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) has a higher predictive power of emotion than do the notes themselves [16]. Clearly the pattern of repetition of chord sequences will depend on the structure and substructures. And the chord sequence lengths will be determined as a product of the syllable counts of the lyrical phrases.

---

[2] e.g., https://wordnet.princeton.edu/

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

[4] e.g., http://www.datamuse.com/api/

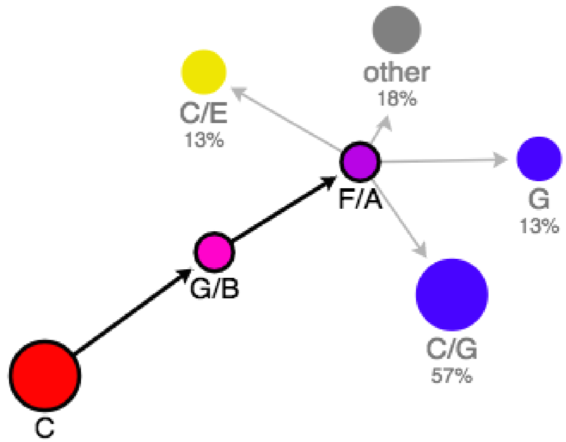


Figure 7: A graphical representation of a Markov model as learned and rendered by Hooktheory.com. In this representation, a context (of length 3) has been selected and the transition probabilities to the next chord are shown.

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.

Unique chord sequences will be generated, consistent with the constraints imposed by the substructure. Chord sequences will be generated in the order in which they are first incorporated into the song so as to ensure at least the presence of a leading context. This is important because the final chord subsequence of a particular segment influences the starting chord sequence of the segment which follows.

The simplest approach to creating harmony would be 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.

The next simplest approach would be to learn phrases of chord progressions from data, which might then be sampled at random to fill substructural elements.

Another simple solution would be 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 [28]. Similar models have been created for pop music and are available via online APIs[5] (see Figure 7). 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, $n^m$ 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

[5] http://www.hooktheory.com/trends

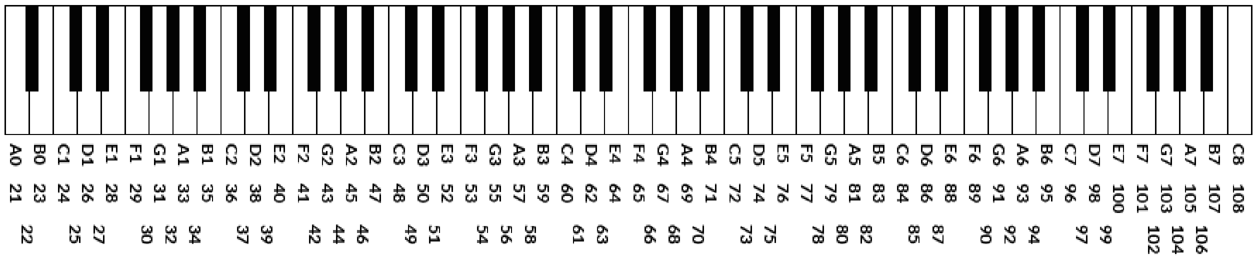


Figure 8: A standard 88-note keyboard labeled with the corresponding MIDI note numbers.

used in [27].

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 would propose that such data is yet sufficient for our purposes inasmuch 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 5 contains a number of chords that we would consider to have been incorrectly transcribed.

It is commonly noted that pop songs suffer from an over-repetition of certain chord subsequences (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 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.

A chord progression (without melody) is not considered plagiarized, even if copied for an entire song. We consider the generation of chord sequences too interesting a problem to take a copycat approach. However, in the interest of generating previously unseen harmonic progressions, adding the non-plagiaristic constraint to harmony generation would be an interesting addition to the model. Plagiarism in Markov sequence generation is considered in [22].

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 (see Figure 8), 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. We intend to do likewise.

The sequence of note (and rest) durations is what is defined as *rhythm*. The alignment of notes with the principle beats of the measure tends towards an emphasis of the corresponding syllables.

A simple 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). Alternatively, durations might be sampled 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 common approach is modeled by [18] 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 a third approach which starts with the lyrics and uses the stress patterns in the lyrics to generate the rhythm.

The generation of pitches to form interesting melodies has been the subject of significant research. A naive approach to generating pitches simply chooses random pitches from those belonging to the triads of the accompanying chords. Markov chains have also been extensively used to address this task.

Once rhythm is determined, pitch will then be generated, not note by note, but by first defining a sequence of idiomatic expressions similar to how Keller defined idiomatic chord progression bricks or how Pachet defined PACTs [14, 20]. 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.

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

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 might be selected so as to maximize singability within a determined vocal range.

Most effective compositions combine the melody and lyrics in such a way as to achieve an effect that is greater than the sum of their 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 "it's all about the *bass*," "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 in 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 (e.g., see [23]), 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 decisions of what voices will be included 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 [23].

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 leadsheet 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.

### 4.7 Self-Evaluation

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 novelty, surprise, and value at each level of the generative system. Each module will thus be able to generate several artifacts for evaluation by the corresponding fitness function in order to determine a single best choice for the compositional element.

## 5 Validation

Ultimately we are interested not in how well we solve each of these subproblems individually, but in our ability to generate a system that expresses musicality that would be deemed creative if expressed by a human. To this end, much of our focus will be on the perceived creativity of the generated artifacts.

Validation in computationally creative research differs significantly from validation in Artificial Intelligence. Whereas AI generally describes well-defined problems with optimal outcomes against which AI solutions can be ranked and or adapted, no ideal outcome exists for computationally creative endeavors.

To characterize creative artifacts, Boden suggests three attributes that have been widely adopted in the validation of computational creative: novelty, surprise, and value [2].

*Novelty* is "the property of an artefact (abstract or concrete) output by a creative system which arises from prior non-existence of like or identical artefacts in the context in which the artefact is produced" [29]. A distance metric is generally used to compute the degree of difference between a newly-generated artifact and all artifacts in the domain space. This metric can then be used to bias towards artifacts whose computed distance is above some minimum threshold (which establishes some minimum definition of novelty) and below a maximum threshold (beyond which the artifact is considered out-of-domain).

Surprise can refer to the unlikeliness, the interestingness, or the astonishment which an artifact exhibits [2]. Meyer is commonly cited for his work in meaning in music and information theory [15]. In essence, when generating a musical sequence, the expectation can be computed (using an information theoretic approach) for any of the possible successive symbols. Generations with high expectation are likely to be good melodies, but are in some sense less interesting. Controlled generation of less expected sequences is one way to introduce surprise to the listener. Surprise then might be considered knowing the "rules" of the art and intelligently breaking them.

As Boden notes, "many arguments about creativity are rooted in disagreements about *value*" [2] (emphasis added). Defining value is more difficult than defining either novelty or surprise, but is nonetheless essential to establishing creativity. Often value is ascribed not only based on the artifact itself, but also based on the process that brought it about [6]. Bias has been commonly been observed against the ability of computer systems to be creative [6]. For this reason, value is commonly evaluated in CC research through the use of human surveys which may or may not reveal that the generative system is a computer.

We plan to use a combination of these approaches to assess the novelty, surprise, and value of each component individually (i.e., structure, lyrics, chords, melody) as well as the system as a whole. In order to assess an individual component, we will control the other variable components by assigning them values taken from existing (less well-known) pop music (e.g., as done in [18]). This way we can directly compare various modular solutions for each component against real-world solutions and possibly against other computationally creative systems which attempt to address the same task while attempting to avoid bias.

## 6 Thesis Schedule

- November 2016 - Get formatted tab data
- January 2016 - Paper on structure submitted to X
- April 2016 - Paper on chords submitted to X
- September 2016 - Paper on melody submitted to X
- January 2017 - Paper on lyrics submitted to X
- January 2017 - Submit dissertation to committee
- March 2017 - Dissertation defense
- April 2017 - Paper on inspiration submitted to X

## 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 *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, 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, Geraint A Wiggins, et al. Computational creativity: the final frontier? In *ECAI*, 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, 1978.

[10] Kemal Ebcioğlu. An expert system for harmonizing four-part chorales. *Computer Music Journal*, pages 43–51, 1988.

[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. Citeseer, 2012.

[12] Paul Ekman. Facial expression and emotion. *American psychologist*, 48(4):384–392, 1993.

[13] Philip N Johnson-Laird. Jazz improvisation: A theory at the computational level. *Representing musical structure, London*, pages 291–325, 1991.

[14] 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.

[15] Leonard B Meyer. *Emotion and meaning in music*. University of chicago Press, 2008.

[16] Rada Mihalcea and Carlo Strapparava. Lyrics, music, and emotions. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 590–599. Association for Computational Linguistics, 2012.

[17] 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. Citeseer, 2010.

[18] 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.

[19] 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.

[20] François Pachet. Representing knowledge used by jazz musicians. In *Proceedings of the International Computer Music Conference*, pages 285–285. INTERNATIONAL COMPUTER MUSIC ACCOCIATION, 1991.

[21] 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.

[22] Alexandre Papadopoulos, Pierre Roy, and François Pachet. Avoiding plagiarism in markov sequence generation. In *Twenty-Eighth AAAI Conference on Artificial Intelligence*, 2014.

[23] George Papadopoulos and Geraint Wiggins. AI methods for algorithmic composition: A survey, a critical view and future prospects. In *AISB Symposium on Musical Creativity*, pages 110–117. Edinburgh, UK, 1999.

[24] W Gerrod Parrott. *Emotions in social psychology: Essential readings*. Psychology Press, 2001.

[25] Geber Ramalho and Jean-Gabriel Ganascia. Simulating creativity in jazz performance. In *AAAI*, volume 94, pages 108–113, 1994.

[26] Marco Scirea, Gabriella AB Barros, Noor Shaker, and Julian Togelius. SMUG: Scientific music generator. In *Proceedings of the Sixth International Conference on Computational Creativity June*, page 204, 2015.

[27] Mark J Steedman. A generative grammar for jazz chord sequences. *Music Perception*, pages 52–77, 1984.

[28] Jukka M Toivanen, Hannu Toivonen, Alessandro Valitutti, et al. Automatical composition of lyrical songs. In *The Fourth International Conference on Computational Creativity*, 2013.

[29] 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 | Has A. Thought |
| Date | Wants A. Change |
| Date | Attends A. Meeting |
| Date | Will B. Present |
| Date | Quinn Snell<br>Graduate Coordinator |