

A funny thing happened on the way to the formula: Algorithmic composition for musical theatre

Nick Collins

nick.collins@durham.ac.uk

Department of Music, Durham University, Palace Green, Durham, DH1 3RL

<http://www.composerprogrammer.com>

TO APPEAR: Computer Music Journal 40(3). This is the last author controlled version submitted. It does not contain responses to copy editing queries and should be considered a preprint but not definitive. In particular, there have been some changes to figures and tables for the final publication not shown here. MIT Press will allow posting of the final version of the article six months after publication.

Abstract

Algorithmic composition methods must prove themselves within real world musical contexts to more firmly solidify their adoption in musical practice. The present project is an automatic composing program trained on a corpus of musical theatre songs to create novel material, directly generating a scored leadsheet of vocal melody and chords. The program can also output based upon phonetic analysis of user-provided lyrics. Chance to undertake the research arose from a television documentary funded by Sky Arts, which considered the question of whether current generation computationally creative methods could devise a new musical theatre work (the research described here provides but one strand within that project). Allied with the documentary, the resultant musical had a two week West End run in London and was itself broadcast in full; evaluation of the project included both

design feedback from a musical theatre composer team, and critical feedback from audiences and media coverage. The research challenges of the real world context are discussed, with respect to the compromises necessary to get such a project to the stage.

Introduction

Academic algorithmic composition projects treating popular music are historically rarer than those investigating such domains as species counterpoint or bebop jazz, though there is a new wave of contemporary activity, perhaps best exemplified by algorithmic methods for electronic dance music (Eigenfeldt and Pasquier 2013; Collins and McLean 2014). The earliest computer music research in automatic composition includes the 1956 pop song generation of *Push Button Bertha* (Ames 1987), or nursery rhyme generation based on information theory (Pinkerton 1956). Yet the predominant investigative domain, as exemplified by the careers of those most famous of algorithmic composers Lejaren Hiller and David Cope, has been classical art music, and in research terms, published work is often restricted to classical training exercises such as chorale harmonization. Opposing this trend, Ames and Domino's (1992) *Cybernetic Composer* was a museum project for a Kurzweil synthesizer able to generate within four popular music styles. More recent manifestations of algorithmic composition within popular culture frequently incorporate interactive control. Where the 1990s saw the *Koan* software and Brian Eno's spearheading of the promotion of generative music (Eno 1996), more recent

manifestations from these authors include the mobile apps *Noatiki* and *Bloom*. Algorithmic procedures have become more visible within digital audio workstations, such as *Max for Live* projects or *Logic's MIDI Scripter*, and appear as the basis of the JukeDeck startup company (jukedeck.com) aiming to provide royalty free generative music for the masses. Such recent work, in the domain of bedroom enthusiasts and corporations as much as academics, has not received much attention in terms of published studies.

Even acknowledging a gathering research impetus into algorithmically generated popular music, prior work on the automatic creation of musical theatre is non-existent. The absence of previous work in automatic generation of musical theatre may be down to a critical rejection of the area as supposedly lacking academic kudos, and a lack of opportunity to get involved with real productions (which are rather high budget enterprises). The present project was motivated by involvement in the Sky Arts funded TV documentary series *Computer Says Show* (Wingspan Productions, 2016), whose premise was the research question of whether computational methods could devise a successful stage musical. Teams of academics (Colton et al. 2016) analyzed existing musicals in terms of setting, plot and audience emotional response, considered automatic book and lyrics generation, audio analysis of cast recordings through Music Information Retrieval (MIR), and in the present case, symbolic composition of song leadsheets. The enclosing project provided real world constraints and deadlines, and promised the ultimate test of a real theatrical West End run.

This article describes the core algorithms for lead sheet generation, both for generating pure song material, and when further constrained to set lyrics. In terms of Pearce, Meredith and Wiggins' (2002) taxonomy, this is computational modeling of musical style, to stand or fall by critical reception; evaluation included within design cycle feedback from the close involvement of a musical theatre director and

composers and TV production staff, and eventually critics and audiences for the real production run. Working towards the ecologically valid final show compromised purity of evaluation that might otherwise have been found in more controlled (and contrived) laboratory circumstances, and raises methodological issues in reaching beyond pure computer music research. It was, however, too good an opportunity to miss, revealing alternative public perspectives on musical algorithms; this article has a further contribution as a cautionary tale for researchers who follow in moving out of the safety of the laboratory.

The leadsheet generation algorithm and its parameters

The software rests upon both corpus analysis of existing musical theatre material, and hard coded rules providing generative constraints, thus combining corpus-based and rules-based work. Corpus work included an automatic chord detection analysis of a large set of musical theatre cast recordings informing a harmony generation model, and a custom corpus of musical theatre song in a novel format which favored analysis, and thus subsequent synthesis, of musical phrases. Phrase materials were subject to Markovian modeling, and analysis statistics also fed into particular production rules. Refinement of the algorithms was chiefly motivated by feedback from the primary documentary participants, two music theatre specialists, Benjamin Till and Nathan Taylor. This process was seen as necessary to constrain the domain of permissible generation to favor a higher proportion of effective outputs. Up-front representational and modeling decisions required in application of machine learning to any corpus are themselves hard coded impositions by the systems designer, and so taking a pragmatic middle way utilizing both corpus- and rules-based techniques was not seen as compromising the project's research.

The code was written in SuperCollider, generating fomis score format text files (Psenicka 2009) as well as parallel MIDI files; MIDI files could be imported in Sibelius, and the fomis software acted as interface to automatic final PDF score generation within Lilypond (MIDI and PDF files were supplied for each leadsheet). Additional external callouts for the lyrics analysis were made to python and the NLTK library (Bird, Loper and Klein 2009). In order to give a taste of the generativity of the software, multiple score examples are given at points below, though such illustrations still remain snapshots of the true large output space.

Chord sequence model

A parallel project, undertaken by Bob L. Sturm, Tillman Weyde and Daniel Wolff, applied MIR analysis to a large corpus of musical theatre cast recordings (from *A Chorus Line* to *Wicked*); the most reliable features for the purposes of training up an algorithmic composition system were provided by chord detection. Chords were extracted throughout using the Chordino plugin (Mauch and Dixon 2010). 53 shows had been marked as 'hits' in an analysis of economic and critical factors by James Robert Lloyd, Alex Davies and David Spiegelhalter (Colton et al. 2016) leading to 1124 analysed audio files totaling around 53 hours of audio.

The chord data is not absolutely reliable, in that the plug-in itself is not as good a listener as an expert musicologist, but does provide a large data source otherwise unobtainable with the human resources to hand. A parsing program was written to translate the textual chord shorthand provided by the Chordino plugin to pitch class note information. Data was cleaned up by removing any 'too fast' chord changes (e.g. quicker than a half a second corresponding to one beat at 120bpm), and ignoring any "N" results where no chord had been found in a given section of audio

(sequences of chords were only considered complete when at least three chords were detected in a row and no "N" intervened).

Having obtained a large set of chord sequences representing hit musical theatre, two chord generators were obtained. In the first case, no attempt was made to impose a home key. In the second, only relative motion between chords fitting within a single major or minor key was permitted to train the model; separate major and minor key models were created. The machine learning algorithm was a prediction by partial match (PPM) variable order Markov model (up to order 3) (Pearce and Wiggins 2004); its application requires integers, so an encoding from chords to integers was created, where ten chord types and twelve chromatic pitches translate to one of 120 possible integers. Figure 1 provides three example generated chord sequences of 24 chords in C major and in C minor, created with the major and minor models, and constrained to start with the root home key chord. Certain loops are evident in the statistics of chord transition; for example, the third minor example includes a case of major to minor chord alteration (on Ab) temporarily stuck in repetition. Chord types are sometimes altered, for example, from a major chord on a particular root to a major chord with added sixth on the same root, potentially lifted from a harmonic sequence or vamping pattern in source material. The chord sequences are generally musical and in character with musical theatre, though without any innovative individual style.

[C, G, G6, F6, Am7, Cmaj7, G, G, Dm, G, C, Cmaj7, Am7, G7, Cmaj7, F, Dm, Em, G, C, G, Fmaj7, Em, C]

[C, G, C, F, G, Cmaj7, C, Em, F, G, C, G, C, G7, C, G7, C, Am7, G, C, G6, Am, Em, F]

[C, F6, Bm7b5, C, F6, C, F, C, F, C, F, G, C, G, C, G, F6, F, G, F, Cmaj7, F, C, C]

[Cm, Bdim7, Cm, Bdim7, Cm, Bdim7, Cm, Bdim7, Ab, Ab, Abmaj7, Cm, Bdim7, Ab6, Cm, Bdim7, Ab, G7, Cm, Fm, Cm, G7, Cm, Fm]

[Cm, Fm, Cm, Ab, G7, Cm, Ab6, Fm6, G7, Baug, G7, Cm, Baug, G7, Cm, Ab6, Ab, G, Bdim7, G, Abmaj7, G, Abm, G]

[Cm, Bdim7, Fm, G, Cm, Ab, Abm, Abm6, Abmaj7, Abm, Bdim7, G7, Cm, G7, Cm, Ab6, Bdim7, G7, Ab, Ab6, Bdim7, Fm6, G7, Baug]

Figure 1: Six example generated chord sequences of twenty four chords, the first three in C major home key (major key chord transition model), the second three in C minor (minor key chord transition model)

A further chord model was obtained by taking the chord transition table data from Declercq and Temperley (2011), which corresponds to a corpus of 100 successful popular music chart songs. Nonetheless, this model was eventually not used for the musical as lacking the specificity of the musical theatre, though it provided a useful comparator.

Melody corpus representation and analysis

Though some musical theatre MIDI files are available online, the reliability and consistency of the data is too variable for immediate corpus work (files are often created by amateur enthusiasts, without any standard track arrangement and often as non-quantized renditions). Since song creation in a passable musical theatre style was the most essential compositional task, requiring stylistically appropriate vocal melody at core, the decision was taken to encode a central corpus of musical theatre

songs as prime exemplars for system training. The encoding fundamentally respected musical phrasing, marking up all melodic phrases explicitly, so as to have an innately vocal melody centered corpus. The two musical theatre experts allied with the documentary team advised on a subset of songs to encode from musicals which had been denoted 'hits' (these musicals included such well known shows as *Cats*, *The Lion King* and *The Rocky Horror Show*).

The encoding provides for a given core song melody its notes as pitch and rhythm, broken down into phrases, associated chords, and a formal denotation of the melody's internal phrase relationships. The melodic data has a redundancy, in that the start and end position of each phrase within a measure, as well as inter-phrase intervals are supplied, but these provide a useful check on human error in encoding (start beat + sum of durations within the phrase should lead modulo time signature measure length to the end beat, which adding the inter-phrase time interval again should lead to the next start beat). An example is in Figure 2, the encoding being itself valid SuperCollider code of nested arrays; the reader can observe the phrase structure with one phrase per line. All melodies were transposed to a home key of C major or minor, and the standard time signature was 4/4, though other time signatures were permissible, and quarter note or half note triplets encodable via beat durations within a tolerance near 0.33 or 0.66. Since representational decisions are key to machine learning, Figure 2 provides insight into the core priorities in musical data for the algorithmic generation system.


```

[
//melody by phrases in form [startbeat within bar (allowing for anacrusis
or initial rest), alternating array of pitch then duration of each note,
end beat of bar of phrase, gap till next phrase]

[
[0,[4,1,-5,1,2,1,-5,1],4,0],
[0,[0,0.5,2,0.5,4,0.5,5,0.5,2,1,7,1],4,0],
[0,[4,1,-5,1,2,1,-5,1],4,0],
[0,[0,0.5,2,0.5,4,0.5,5,0.5,2,1,7,1],4,0],
[0,[9,0.5,12,0.5,12,0.5,12,0.5,14,1,12,0.5,11,0.5],4,0],
[0,[9,0.5,12,0.5,12,0.5,12,0.5,14,1,12,0.5,11,0.5],4,0],
[0,[9,0.5,12,0.5,12,0.5,12,0.5,14,0.5,12,0.5,9,0.5,5,0.5,7,2],2,1.5],
[3.5,[4,0.5,2,0.5,2,0.5,2,0.5,4,0.5,5,0.5,7,0.5,4,0.5,2,0.5,0,2],2,2]
],
//chord sequence, as alternating array of pitches of the chord, and
associated duration

[ [0,4,7],2,[0,4,7]+7,2,[0,4,7],2,[0,4,7]+7,2,[0,4,7],2,[0,4,7]+7,2,[0,4,7
],2,[0,4,7]+7,2, [0,4,7]+5,2,[0,4,7],2,[0,4,7]+5,2,[0,4,7],2,[0,4,7]+5,2,[
0,4,7]+10,2,[0,4,7],4,[0,4,7]+7,4],

//medium scale form, inter-relationship of phrases, in this case
ABABCCCD

[0,1,0,1,2,2,2,3],
]

```

Figure 2: Example encoding of Andrew Lloyd Webber's *Music of the Night*, from *Phantom of the Opera* (1986), with annotated comments

45 songs were encoded in this manner; encoding was a relatively intensive process, requiring analytical decisions on phrase boundaries and phrase relationships that may be points of disagreement between analysts, but which were of sufficient quality to form the basis for novel generation of phrases.

The phrase based encoding allows for statistical analysis of a number of attributes of phrasing in musical theatre material. As would be expected from music psychology, phrase durations (assuming an average tempo of 120 bpm) were around 3 seconds in length, corresponding well to the perceptual present and associated memory constraints (London 2012). Chromatic movement was much rarer than diatonic (2052 diatonic note transitions compared to 213 chromatic), as might have been anticipated for popular music theatre melody. Note to note pitch interval movements were more frequently by step than by leap (that is, larger than a proximate step), in the proportions 44.66% (adjacent step) 23.26% (same note) 16.68% (leap up) 15.4% (leap down). Of 604 leap intervals, 216 were followed by a step, 214 by another leap (65.9% of the time in the opposite direction to the previous leap) and 174 were the last interval in a phrase.

Statistics were also extracted for phrase ranges, including mean and median phrase pitches. A whole transcribed song extract could provide guide templates for melodic movement. Melodic corpus phrase data provided the basis for variable order Markov models over pitches, melodic intervals, contour classes, durations and inter-onset interval classes useful for novel melody generation founded in corpus statistics. Assuming 4/4 (the majority of the melodies conforming to this time signature), statistics were also obtained on pitch choices and pitch intervals at each distinct eighth note of the measure.

Melody generation algorithm

The melody generation algorithm creates musical materials at a local level of the phrase, with a medium scale structure built up by the phrase inter-relationships to create song sections, and the final song built up by repetition of sections within a form. The phrases of the melodies in the training corpus are used to train pitch and rhythm models, to construct novel phrases. Novel phrases are specified within a diatonic pitch space, and in their re-use these phrase materials are thereby automatically adjusted to work against changing harmonic contexts. The source melodies also provide guidelines for the form over multiple phrases, including the skeleton of pitch height over a melody. The idea of guide melody mean pitches constraining new generation bears a relation to the use of an elastic tendency towards the mean pitch of the phrase within previous psychologically inspired treatments (Brown et al. 2015).

The central melody generation routine has quite a number of control arguments, listed in Table 1, giving insight into the flexibility of the calculation. In a number of places, corpus-derived models and statistics naturally inform the underlying routine.

Argument	Result	Default
Key	Set base key for generation	C major
Time signature	Set base time signature; no compound signatures, typically 4/4 or 3/4	4/4
Range	Set singer's range, permissible compass of notes	0 to 12, one octave

Chords	Chord sequence to work to (from chord model, or imposed)	Generated from chord model
Eighth note data	If true, utilize statistics separately collated for each eighth note of the bar, rather than aggregated across all positions	50% / 50% true / false
On beat chord probability	Probability of restricting on beat positions to only use notes of the current chord	100%
Allow sixteenth notes	Allow faster rhythmic units within a melody	100%
Pitch choice model	Select between two available pitch choice models, one based on a greedy dynamic programming approach, and one a variable order Markov model	Greedy dynamic programming
Top jump	Top leap size in diatonic steps	8
Patter rhythm probability	Chance of rhythm generation using a 'patter rhythm', that is, fast sequence of durations as per Gilbert and Sullivan's <i>I Am the Very Model of a Modern Major-General</i>	0%
Use PPM for rhythm	Whether to use a prediction by partial match model for generating rhythmic sequences, or a rule based process	0%
Max contiguous syncopation	Maximum number of notes which can be syncopated (falling on an off-beat) in a row	2
Guide	Whether a template phrase pattern can influence	100%

strictness	pitch position (the guide consists of the average pitch per phrase)	
Impose form	User specified phrase form rather than derived from a guide melody	false
First chord is tonic	Enforces any generated chord sequence to begin on tonic chord of the key	false

Table 1: Control arguments for the central melody generation function

Figure 3 presents two example leadsheets, each restricted to eight measures only, to give a flavor of the generation. The parameters are the defaults for the leadsheet generation algorithm as per the last column in the table. No attempt has been made to cherry pick, these being the first two created directly for this example.

example 3.1

example 3.2

Figure 3: Two example generated lead sheets of eight bars

Ostinato generation algorithm

A frequent requirement for musical theatre composition is the creation of rhythmic and pitch ostinati, as backings during songs and instrumental filler music, with a strong connection to popular music styles. Similar principles to the vocal melody generation work were applied, but with a separate corpus consisting of some well known ostinato from popular music and musical theatre (e.g., Michael Jackson's *Smooth Criminal*, Queen's *Another One Bites the Dust*, *One Day More* from *Les Misérables*).

The backing harmony was either C minor or C major, with no other chord changes; the expected use was that the ostinato could be adjusted to match other chords in a song if needed, but was in its most basic manifestation for a groove on a set root. Figure 4 provides a variety of example outputs (again, the first set generated for this figure). Note the overly wide ranging movement in the seventh, the common initial rhythmic pattern in the first and third, and the appearance of dotted and Scotch snap rhythms in the C minor patterns, as well as the syncopation of the sixth ostinato.

The image displays eight musical staves, each representing a different ostinato pattern. The staves are arranged in four pairs, with each pair corresponding to a specific key signature: C major and C minor. Each staff begins with a treble clef and a 4/4 time signature. The first staff is labeled 'C' and features a sequence of eighth notes: G4, A4, B4, C5, B4, A4, G4, with a final quarter rest. The second staff is labeled 'Cm' and features a sequence of eighth notes: F4, G4, A4, Bb4, A4, G4, F4, with a final quarter rest. The third staff is labeled 'C' and features a sequence of eighth notes: G4, A4, B4, C5, B4, A4, G4, with a final quarter rest. The fourth staff is labeled 'Cm' and features a sequence of eighth notes: F4, G4, A4, Bb4, A4, G4, F4, with a final quarter rest. The fifth staff is labeled 'C' and features a sequence of eighth notes: G4, A4, B4, C5, B4, A4, G4, with a final quarter rest. The sixth staff is labeled 'Cm' and features a sequence of eighth notes: F4, G4, A4, Bb4, A4, G4, F4, with a final quarter rest. The seventh staff is labeled 'C' and features a sequence of eighth notes: G4, A4, B4, C5, B4, A4, G4, with a final quarter rest. The eighth staff is labeled 'Cm' and features a sequence of eighth notes: F4, G4, A4, Bb4, A4, G4, F4, with a final quarter rest.

Figure 4: Eight generated ostinati (four examples each for C major and C minor).

Generation based on lyrics

Musical theatre composition can proceed led by a musical idea first, or from a lyric. In order to accommodate a frequent request of the show developers to accommodate existing text, a front end process was devised to analyze song lyrics and be able to set notes to their implicit accent pattern.

Code utilized the Python library NLTK (Bird, Loper and Klein 2009), which provides a function to analyze metrical stress within a word over syllables, as well as a dictionary from the Gutenberg organization (<http://www.gutenberg.org/files/3204/>) which provided exact syllable breakdowns for common words (e.g., "ac-com-mo-dat-ing", "un-cal-cu-lat-ing"). Text was provided as a block, converted to lower case ascii without special characters, and separated by line (using newlines) and words (using spaces). The prepared text was fed to an external python program (passing data to and from SuperCollider via auxiliary text files), where the metrical stress analysis came down to a special dictionary lookup (in the `cmudict.dict()` available with NLTK, which supplies per word analyses). The python library gives stresses at three levels, for example, for the text below:

```
"i got extremely bored of the never ending discussion of authorship around
```

```
1 1 0 1 0 1 1 0 1 0 1 0 0 1 0 1 1 0 2 01
```

```
generative art" (Alex McLean, from a facebook post)
```

```
1 0 0 0 1
```

"authorship" is marked 102 so that "ship" is the highest stress in the whole sentence.

Musically, a reconciliation must be effected between the stress pattern and the metrical frame provided by the time signature; good scansion would normally indicate strong stresses of syllables on strong beats. Syllables (all of which have an associated vowel for singing) might be extended via melisma, but that option was not pursued in the current case. Instead, syllables were allocated measure position based on a default of offbeats for stress level '0', and on-beats for '1'; in 4/4, a succession of '0's could fill in across eighth notes, but successive '1's would be spaced by quarter note beats.

Figure 5 provides three examples generated using Alex McLean's text. In all three, the split of "end-ing" with "end" on a quarter note shows the lack of flexibility of the software to certain possibilities of patter (end-ing could be two eighth notes in line with other parts of that phrase). Note how "ship" always falls to an on-beat.

Figure 5 shows three example generated lead sheets from lyrics. The lyrics are: "i got ex treme ly bored of the nev er end ing dis cus sion of au thor ship a round gen er a tive art". The chords are: G, Cmaj7, G, Dm, Am, G6, Bm7b5, F6, G7, Cmaj7, G6, F6, Em, G7, Am7, G, Am7, G.

Figure 5: three example generated lead sheets from lyrics

The algorithm presented here has trouble with lyrics with a strongly repeating line by line pattern, denoting a common anacrusis, and favors 4/4 over 6/8 interpretations. A facility was added to force a particular pick up structure on the output. It proved practical for generation for this project, but would be open to much future improvement; the natural language dictionaries themselves were also found to be rather incomplete for song lyrics. In some cases, words had to be provided split up ahead of the syllablization process (the dictionaries might be extended themselves to solve this).

This form of text to music generation is in contrast to (but might be expanded through) sentiment analysis based work, such as the wonderfully named *TransProse*

system (Davis and Mohammad 2014), which creates piano pieces based upon the emotional subtext of novels. There is little prior work generating songs directly from lyrics, excepting for systems such as *MySong/Songsmith* (Simon, Morris and Basu 2008), or app developer Smule's mobile software *Songify* and *AutoRap*, which operate by onset and pitch detection within the audio signal and carry text along with them.

The human-computer collaborative design of the final music

A fully autonomous complete musical theatre leadsheet generating program was created, combining the melody generation and chord generation modules, coupled with some rules on form. In practice, however, operation of the program was in the domain of computer-assisted composition (Miranda 2009), used to provide material that was then manipulated by human composers. The compromises of working within a high profile broadcast project with multiple stakeholders necessitated more human intervention before public performance than would have been preferred for pure research; but then, access to a West End venue for evaluation would never have occurred without such oversight.

To maintain some researcher objectivity concerning aesthetic choice at the heart of song selection, batches of computer generated outputs were sent en masse (often one hundred songs at a time), without any cherry picking, to the musical theatre specialists. The human composition team essentially selected fragments (somewhat laboriously and without consultation with the research team) from 607 song lead sheets and 1171 ostinati, working with a rehearsal pianist. After particular discovery sessions and in the process of musical development of the final musical theatre piece, they sent requests for revisions and novel program output, for example, soliciting a suite of songs in 3/4 instead of 4/4. The musical theatre

composers' musical preferences and narrative needs had an unavoidable influence on the material making it through to the show, and they frequently freely composed around the skeleton of computer generated material. The TV production company had mandated an intention to respect the computer generated material; that the human composers felt able to still range widely from this base is some indication of both limitations in the algorithmic composition, and discomfort in the task of negotiating between algorithm and human vision.

Table 2 lists the 16 songs in the show, and their derivation from the computer programs involved in the production. In some cases, the human composition team has only kept a minimal fragment of melody, or in the worst scenario, just a chord sequence (which is a less than unique data point, uncopyrightable, and trivially taken unrecognizably far from the original generated material). The production team compiled with the human composers a document detailing the origins of each song in the show (Till et al. 2016), so as to track the experiment and to assess authorship proportions with respect to publishing rights; some relevant quotes are reproduced in the table, which uses this source, alongside further analysis of the songs, to attribute the algorithmic components. To complicate matters, the Flow Composer software (Pachet and Roy 2014) was also used to contribute towards a few songs, though it is beyond the scope of the present article to further evaluate that software here (see Colton et al. 2016 for more on the role of Flow Composer).

The final column of Table 2 gives an estimated percentage of computer composed contribution to the final songs for the algorithm presented in this article ("ALW"). The percentage is derived from musical analysis of the final pieces against the source algorithmically composed lead sheets, and from examination of human composer comments on their manipulation of the source song material (Till et al. 2016). This calculation was necessitated by UK Performing Rights Society registration for the musical, which forced a quantitative decision. The overall

average contribution for the computer over the 15 songs where ALW was utilized works out as 32%, or around one third of the composition. Whilst this number cannot be seen as definitive, given the limitations of human self-reflection on creative acts and the working opacity of the machine algorithm, it is suggestive of the process. In cases where two human composers were intimately involved in songs, it points to an equal three way split between authors (two humans and a computer); however, in many cases a single human composer worked on a given song, and the contribution percentage is less impressive.

Table 2: Songs in the show and their derivation

Song	Which Algorithm	Algorithmically generated material	Extent of computer involvement	Estimated computer contribution percentage
1. Green Gate	ALW	2 ostinati, chord sequence, melody and chords	Computer composed eight bar theme starts the show, and is basis of much further material	50
2. We Can Do It Alone	ALW	16 bar 3/4 central section chords and melody line	As accompaniment figure in central section, otherwise human composed including singing part over the top	20
3. Penetrate The Base	ALW	Chord sequence and two ostinati	Chord sequence, intact but with interpolated B minor, obvious underneath verse though human	40

			<p>composed lead vocal. Ostinati are used quite strongly in the composition; the main ostinato is slightly adjusted from the computer original through its derivation is clear, the second appears later in the song. "I hope the use of this ostinato through this number and at other key dramatic moments of the show will give it the same impact as the ostinato which starts Heaven On Their Minds from Jesus Christ Superstar and is later used for the whipping scene. This was one of the references given [to the researchers]... I feel the creation of ostinati was a very successful aspect of this process because it also allowed me a great deal of creative freedom when working out what was going on around the ostinato." (Till et al. 2016, p. 11)</p>	
4. So Much To Say	ALW	Melody and chords	The middle section melody of the piece can be traced to a few bars	20

			of program output, but otherwise humans had much more to say	
5. Graceful	ALW	Melody and chords generated to lyrics	Possibly the most substantially respected computer generation, though there is certainly tweaking of output to best fit lyrics where the automated scansion fails, and additional human composed material.	50
6. We Are Greenham	Flow Composer	Lead sheet created based on Greenham protest songs	Quite well respected, see Colton et al. 2016.	N/A
7. At Our Feet	ALW	Melody and chords	Much of the material is highly related to the computer source. Core catchy elements in verse and chorus are indicated by the computer part, though have been rhythmically tweaked (to the better) by human hand.	50
8. Unbreakable	ALW and Flow Composer	ALW: melody and chords Flow Composer: melody and chords	Shows some connection to computer original materials, though human tweaking especially in shifting to a calypso style	30
9. How	ALW	Melody and	A single leadsheet led to all the	50

Dare You		chords	<p>source materials for the song; some rhythms have been changed, in particular from straight half notes to less symmetrical quarter and dotted half, but the main verse is a clear derivation from the computer. The chorus is a greater stretch to relate, though has a basic intervallic cell in common, if shifted in rhythm. Setting to lyrics led to more elaborate human composed melodic variations.</p>	
10. Bouncing Back	ALW	Melody and chords generated to lyrics	<p>The computer output was substantially adjusted in rhythm because of the demands of the lyrics, and failings in its appreciation of natural scansion "... as a comedy song, the rhythms of the lyrics are so important for the comedy aspect. Break the rhythm that is inherently in the words, and you lose so much of the comedy. As we know already, this system doesn't yet have much of a grasp of stressed syllables vs.</p>	50

			unstressed ones, let alone meter and form, such as dactyls, iambs and spondees!" (Till et al. 2016, p. 32)	
11. Would It Be So Bad	ALW	Melody and chords	The computer source is mainly lost here against human composed material, though is more apparent in the closing ensemble material based on a different lead sheet.	30
12. Scratch That Itch	Flow Composer and ALW	Both programs provided melody and chord material	Much of the computer material was cut in rehearsals, leaving just some fragments of chord sequences of doubtful clear relation to the original	10
13. What's The Point	ALW	Melody and chords	In the main part of the song, only chord sequences from the computer were used with the rest human composed. The middle eight is claimed to rest on a computer composed leadsheet (Till et al. 2016 p. 42), though the relationship is too stretched to be apparent.	10
14. In Our Hearts	ALW	Melody and chords	Corrections were made to the rhythm for improved lyrical	40

		generated to lyrics	setting, but computer material is clearly present in the final version including the main chorus melodic hook.	
15. Thank You	ALW	Melody and chords	The initial trajectory of the song is determined by a 3/4 fragment of computer composition, though the main onrush of the song with its frantic melodic movement bears little relation	30
16. Beyond The Fence / At Our Feet / We Are Greenham / Green Gate	ALW and Flow Composer		The first part of this closing number is another "computer-inspired" (Till et al. 2016, p. 56) treatment, taking one program output song as an initial guide. A recap of various parts of the show follows, though the human hand in the composition remains clear.	25

Figure 6 shows the first four bars of the computer composed chorus material, versus the eventual human doctored show tune for 'At Our Feet'; there is a relation, but there is also a reworking going on that moves rhythms towards more comfortable patterns, streamlines melody, and isn't afraid to reharmonize. The result is a more conventional musical theatre composition, and the nature of these adjustments is

actually of strong potential in showing future revision possibilities for the generating algorithm.

Computer generated original



Human re-interpretation



The figure displays two musical staves. The top staff, titled 'Computer generated original', is in 4/4 time and features a melody with chords Bb, F, Bb, and C7. The bottom staff, titled 'Human re-interpretation', shows a modified melody with chords Bb, F/A, Bb, and F/A. The notation includes notes, rests, and bar lines, with the second staff ending in a double bar line.

Figure 6: Computer generated original chorus material versus eventual human finished song

In many cases in the show, a claimed link between computer composed original and the eventual show score is only vaguely perceptible, or obfuscated by transformations such as rhythmic value substitution, new pitches or chord substitutions, and shifting with respect to barlines to change metrical emphasis (particularly and perhaps forgivably used for instances of generation to lyrics). Orchestration in the final production was carried out entirely by human hand, and the live band at the show provided some inherent ambiguity as to the music's origins (the score featured quite a lot of electric guitar in power rock vein).

Evaluation through critical reaction

Few algorithmic composition projects have had the opportunity to receive critical appraisal in a high pressure real world situation with wider exposure than an art

music concert of cognoscenti. Though, as detailed in the previous section, the material had gone through human modification to varying degrees without the involvement of the original researchers, there was a computational presence within the final musical theatre piece. On 26th February 2016, a real West End theatre show was judged by real theatre critics from national media, and the show had a two week run around this gala performance (Figure 7).



Figure 7: The musical at the Arts Theatre, London

The theatre reported well engaged audiences, with decent attendance over the fortnight run, with many positive twitter comments and other public feedback.

3047 people saw the musical, or around 60% of the theatre's seating capacity during the run (there was virtually no wider marketing budget for the show and attendance generally followed press the algorithmic ideology had attracted). As far as it is possible to poll, audiences were mainly drawn from typical West End musical theatre goers, with an unknown proportion of tech sector workers and academics, who may have attended due to the novelty of the generative component. The press night had a greater proportion of family and friends of cast and creative team. For the final three performances, audiences were polled by Wingspan Productions and asked to rate their enjoyment of the show from 1 (low) to 5 (high). Of 57 respondents, the poll revealed an overwhelmingly high level of enjoyment (1/1.7%, 2/1.7%, 3/10.3%, 4/17.3% and 5/69.0%).

However, theatre critics are a more volatile group. Table 3 accumulates some of the most pertinent critical judgments, with a particular emphasis on comments on the music specifically. The more astute critics, such as The Telegraph's Dominic Cavendish, picked up on the level of human intervention in the final production: "Beyond the Fence has – if nothing else – considerable curiosity value, even if that value diminishes when you find out about its actual genesis. This experiment to see whether state-of-the-art computing might deliver the next Sound of Music has plainly benefited from a lot of human intervention in the six months it has taken to get from its preliminary boot-up to the West End stage. To call it "computer-generated" is misleading. "Computer-initiated" and "computer-assisted", though less grabby, are more accurate" (Cavendish 2016).

The broad consensus was that the underlying show was passable but by no means outstanding. In some ways, this is a success for stylistic composition, though the human cherry picking from and finessing of the raw computer output provides an additional layer of filtering that tempers confidence in a strong result. That the show was not ground breaking in its music is unsurprising given the reliance on

databases of musical theatre music across decades. Statistical analysis aggregated across time periods, simply selecting hit musicals without any concern for recent trends in musical theatre; unsurprisingly, critics picked up on this averaging effect. Design by committee is a lurking issue at the heart of the production.

Table 3: Selected critical reception in media outlets

Outlet	Reference	Rating (out of 5)	Quote
The Stage	Vale 2016	3*	'Little, if any, new ground is broken, either in the structure or the score...a varied score'
The Telegraph	Cavendish 2016	3*	'It might have been more satisfying all the same to plump for a scenario of an ostentatiously technological nature, or at least take inspiration from the "new wave" electronica of the time...It looks and sounds analogue, generic, presses no avant-garde buttons... a terrific end-of-show number [Thank You] ..."Computer says so-so" then. In a world where flops are the norm, no mean feat'
The Independent	Williams 2016	3*	'The result, as you might expect, feels formulaic. The music, piano-led ballads and squealy 80s power-rock, sounds vaguely familiar yet there are no barnstorming, hummable hits...I wonder if the computer-generated tag will help or

			hinder: it's hard to think you'd watch the show without being more interested in the process than the product. And am I being romantic in thinking it's telling that while the story and songs work fine, the thing that makes it zing is the human-chosen setting? Maybe, but I don't think theatre-makers need to start smashing computers any time soon'
The Guardian	Gardner 2016	2*	'a dated middle-of-the-road show full of pleasant middle-of-the-road songs'
Londonist	Black 2016	3*	'It's quite fun to try and spot stuff the tech has re-purposed: a bit of Chicago here, a bit of The Lion King there — quite a bit of it sounds like Meatloaf at medium throttle.'

The project did lead to much media publicity, and can be seen as a landmark in public exposure to computational creativity (Colton et al. 2016). Perhaps the most apt coverage was the New Scientist article which quoted from the biography created for the algorithmic composition program: 'Other interests include composing music for musical theatre, composing musical theatre music, music theatre composition, and the overthrow of humanity' and clearly understood the inchoate technology and its averaging effects: 'For all the algorithmic cleverness behind the technology, a huge amount of its heavy lifting amounts to a kind of fine-grained market

research... the UK's musical theatre talents can sleep peacefully at night with little to fear from ... cybernetic pretenders' (Pringle 2016).

In the course of the research after media coverage, a legal letter was received from a well known musical composer concerned at the use of a parodical version of his name for the program, and seeking to stop this under trademark law. That letter is quoted here under fair dealing for the purposes of critique, illuminating as it is to bias in the old school entertainment establishment and the backwardness of the law confronting new computational possibilities:

'In addition, our client is concerned about the imputation which is carried by naming the Program 'Android Lloyd Webber'. Our client is an innovative composer, yet the name of the Program can be understood to imply that our client's musicals have been composed by way of a mechanical process rather than a creative process, which is derogatory.' (Ashby 2016)

It seems more derogatory that a 'mechanical' (computer programmed) process could not be creative, especially in terms of the creativity of the human author of such a program. It also seems a contradiction to seek to stop a program on commercial grounds from producing output that could be confused with that of a human, and at the same time be so worried as to denigrate the program's capabilities in emulating creativity.

Figure 8 provides a gentle response to criticisms by setting selected comments in a song. This is the first pure output of the program, untouched by further human composition; some motivic reuse is clear, though the melodic line doesn't stray far. As presented in bare score, there is no human performance mediation; the songs for the musical had the benefit in performance of human expression, and human editing and orchestration. These provide a further confound to experimental control, though again we must offset this problem against the ecological validity of the final product.

The image shows a musical score in 4/4 time, written in treble clef with a key signature of one flat (Bb). The score consists of five staves of music, each with lyrics underneath. Chord symbols are placed above the notes. The lyrics are: "composed by way of a mechanical process rather than a creative process quite a bit of it sounds like meat loaf at medium throttle vaguely familiar yet there are no barnstorming hummable hits presses no avant garde but tons little if any new ground is broken which is derogatory".

com posed by way of a me chan i cal pro cess ra ther than a cre a tive pro cess
 5 quite a bit of it sounds like meat loaf at me di um throt tle
 9 vague ly fa mil i ar yet there are no barn storm ing hum ma ble hits press es
 13 no a vant garde but tons lit tle if
 16 an y new ground is bro ken which is de rog a to ry

Figure 8: Some critical reaction algorithmically set to song

Three recommendations are gathered here for future algorithmic composers, that is, those who create algorithmic composition programs, in the position of working with a musical theatre team:

- 1) Expect a push from the musical theatre specialists for heavy post algorithm human editing, and try to stay involved in later stages of the production process
- 2) It may be more productive, given the current close links of musical theatre composition to popular music, to create an effective pop song generator with clearly demarcated verses and choruses, and some step-up transpositions of materials, rather than attempt to work against a corpus of many decades of musical theatre shows. For deeper evaluation purposes, a larger historical

corpus of musical theatre shows should be broken up and subsets assessed to ascertain the effect of different eras on output.

- 3) Musical theatre critics may be disappointed that a computer generated musical doesn't engage with computational topics as its essential subject matter. If an algorithmic composer aims to blend in with a mainstream of musical theatre composition, success may be taken as blandness of vision!

Despite these challenges, which should not be underestimated as obstructions to pure computer music research, there are great rewards in a real world project reaching a wider audience beyond specialists. Ultimately, algorithmic composition research must engage with this wider sphere to increase the exposure of such ideas within culture. Since music ultimately stands or falls on general reception, rather than controlled laboratory studies, it is prudent to take opportunities to engage with larger public facing projects, though methodologies will need careful finessing in future research. The hope is that there are essential aspects of the act of human composition to be discovered through such higher profile musical modeling challenges.

Conclusions

Computational music generation towards a West End show provided a rare chance for very public reaction to algorithmic composition. Despite the clear publicity for 'the world's first computer generated musical' the final piece was highly mediated by human intervention, though much of the musical seed material did originate algorithmically. Whilst the demands of an associated television documentary series and human interventions ahead of performance clouded the purity of evaluation, it

has been possible to still discover new facets of practical generative music based on corpora, and explore text-driven creation of leadsheets. These techniques should also be applicable within various domains of popular music generation, in the first instance by switching the source corpus to one of appropriately annotated popular songs. Though methodology necessarily remained pragmatic in negotiation with real world deadlines and output, the present work should serve as a case study and cautionary tale for future projects which seek to move from academia to fully ecologically valid contexts.

Future work might investigate a number of alternative approaches. Cleaned up MIDI files may provide a route to a larger corpus of symbolic material. A historical investigation into musical theatre composition might benefit from an online repository of late 19th and early 20th century works hosted by the Gilbert and Sullivan Archive, with many MIDI files created by Colin M. Johnson in particular (Howarth 2016). A more complicated model of text setting would be crucial to more effective automation of song production, allowing for deliberately extended syllables via melisma, and picking up more effectively on repeated stress patterns over lines indicative of a common anacrusis. Musical theatre composition itself has not been the prime subject of previous algorithmic composition research but deserves wider future investigation, as a site of popular contemporary compositional practice; interaction with traditional human composers has much remaining to teach algorithmic musicians.

Acknowledgements

Catherine Gale designed and led the overall project, ably co-ordinating the various research teams involved, and her vision is ultimately to thank for the project

reaching the stage. Sky Arts were responsible for the majority of the funding, and Wingspan Productions created the TV programmes and led organization of the associated theatrical run and media coverage. Archie Baron as head of Wingspan, and the musical theatre specialists Benjamin Till and Nathan Taylor, gave unstintingly of their time and energy to negotiate through the implications of the algorithmic musical material. Archie Baron also supplied the data on theatrical attendance and the audience poll included in the article. Simone Tarsitani and James Tate assisted in encoding musical theatre songs for the corpus used as a basis for the algorithm. Bob L. Sturm provided feedback on an earlier draft, and three anonymous reviewers provided later suggestions ahead of the final published form. The third reviewer disliked the main title and suggested a few alternatives including 'AIs and Dolls' and 'Random of the Opera'. For the record, a fuller list of puns created for this project includes *Guys an Droids*, *Computer Side Story*, *Bits*, *Phantom of the Musical*, *Boot of Mormon*, *Joseph and his Technocoder Dreamcode*, *Bytety and the Beast*, *Miss Cygon*, *Programmer on the Roof*, *My Fake Lady*, *okcoder!*, *Crazy For You* and *Spamabot*.

References

- Ames, C. and Domino, M. "Cybernetic Composer: an overview." In M. Balaban, K. Ebcioğlu, and O. Laske (eds.) (1992) *Understanding Music with AI: Perspectives on Music Cognition*. Menlo Park, CA: The AAAI Press/MIT Press, pp. 186-205
- Ashby, K. 2016. Personal communication.
- Bird, S., Loper, E. and Klein, E. 2009. *Natural Language Processing with Python*. O'Reilly Media Inc.

- Black, S. 2016. Computer-Penned Musical Beyond The Fence Reviewed. Posted online 29th February 2016. <https://londonist.com/2016/02/computer-penned-musical-beyond-the-fence-reviewed>
- Brown, A., Gifford, T. and Davidson, R. 2015. "Techniques for Generative Melodies Inspired by Music Cognition." *Computer Music Journal* 39(1): 11-26
- Cavendish, D. 2016. Beyond the Fence, Arts Theatre, review: 'computer says so-so'. Posted online 27th February 2016. <http://www.telegraph.co.uk/theatre/what-to-see/beyond-the-fence-arts-theatre-review-computer-says-so-so/>
- Collins, N. and McLean, A. 2014. "Algorave: live performance of algorithmic electronic dance music." *Proceedings of New Interfaces for Musical Expression*, London
- Colton, S., Llano, M.T., Hepworth, R., Charnley, J., Gale, C.V., Baron, A., Pachet, F., Roy, P., Gervas, P., Collins, N., Sturm, B., Weyde, T., Wolff, D. and Lloyd, J. 2016. The Beyond The Fence Musical and Computer Says Show Documentary. *International Conference on Computational Creativity*, Paris.
- Davis, H. and Mohammad, S. M. 2014. "Generating Music from Literature." *Proceedings of the 3rd Workshop on Computational Linguistics for Literature (CLfL) @ EACL 2014*, pp. 1–10, Gothenburg, Sweden, April 27, 2014.
- De Clercq, T. and Temperley, D. 2011. "A corpus analysis of rock harmony." *Popular Music* 30(1): 47-70
- Eno, B. 1996. "Generative Music." *In Motion Magazine*, July 7th 1996. Available online at <http://www.inmotionmagazine.com/eno1.html>
- Eigenfeldt, A. and Pasquier, P. 2013. "Considering vertical and horizontal context in corpus-based generative electronic dance music." In *Proceedings of the Fourth International Conference on Computational Creativity*, Sydney, Australia.
- Gardner, L. 2016. Beyond the Fence review – computer-created show is sweetly bland. Posted online 28th February 2016.

<http://www.theguardian.com/stage/2016/feb/28/beyond-the-fence-review-computer-created-musical-arts-theatre-london>

Howarth, P. (ed.) 2016. *British Musical Theatre*. Online archive.

<https://math.boisestate.edu/gas/british/>

London, J. 2012. *Hearing in Time: Psychological Aspects of Musical Meter* (2nd Edition). New York and Oxford: Oxford University Press

Mauch, M. and Dixon, S. 2010. "Approximate Note Transcription for the Improved Identification of Difficult Chords." *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR 2010)*, Utrecht, Netherlands.

Miranda, E. R. 2009. "Preface: Aesthetic Decisions in Computer-Aided Composition." *Contemporary Music Review* 28(2): 129-132

Pachet, F. and Roy, P., 2014. Imitative Leadsheet Generation with User Constraints. *Proceedings of European Conference on Artificial Intelligence (ECAI)*, Prague, Czech Republic

Pearce, M., Meredith, D. and Wiggins, G. 2002. "Motivations and methodologies for automation of the compositional process." *Musicae Scientiae* 6(2): 119-47

Pearce, M. and Wiggins, G. 2004. "Improved Methods for Statistical Modelling of Monophonic Music." *Journal of New Music Research* 33(4): 367-85

Pinkerton, R.C. 1956. Information theory and melody. *Scientific American* 194: 77-86.

Pringle, S. 2016. Beyond The Fence: how computers spawned a musical. Posted online 3rd March 2016. <https://www.newscientist.com/article/2079483-beyond-the-fence-how-computers-spawned-a-musical/>

Psenicka, D. 2009. Automatic Score Generation with FOMUS. *Proceedings of the International Computer Music Conference*, Montreal.

Simon, I., Morris, D. and Basu, S. 2008. "MySong: automatic accompaniment generation for vocal melodies." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Florence.

Till, B., Taylor, N., Mace, K. and Wingspan Productions Ltd. 2016. *Beyond the Fence annotation notes*. Internal report, Wingspan Productions.

Vale, P. 2016. Beyond the Fence review at Arts Theatre, London – 'futuristic composition with traditional problems'. Posted online 26th February 2016.
<https://www.thestage.co.uk/reviews/2016/beyond-the-fence-review-at-arts-theatre-london-futuristic-composition-with-traditional-problems/>

Williams, H. 2016. Beyond the Fence, Arts Theatre, review: Despite my reservations I was won over. Posted online 27th February 2016.
<http://www.independent.co.uk/arts-entertainment/theatre-dance/reviews/beyond-the-fence-arts-theatre-review-despite-my-reservations-i-was-won-over-a6900836.html>

Wingspan Productions. 2016. "Computer Says Show" (directed by Catherine Gale) and "Beyond the Fence" (Wingspan Theatricals/Sky Arts, television broadcast directed by Tim van Someren). wingspanproductions.com