RESEARCH

I Keep Counting: An Experiment in Human/AI Co-creative Songwriting

Gianluca Micchi*, Louis Bigo*, Mathieu Giraud*, Richard Groult[†,‡] and Florence Levé[‡,*]

Musical co-creativity aims at making humans and computers collaborate to compose music. As an MIR team in computational musicology, we experimented with co-creativity when writing our entry to the "AI Song Contest 2020". Artificial intelligence was used to generate the song's structure, harmony, lyrics, and hook melody independently and as a basis for human composition. It was a challenge from both the creative and the technical point of view: in a very short time-frame, the team had to adapt its own simple models, or experiment with existing ones, to a related yet still unfamiliar task, music generation through AI. The song we propose is called "I Keep Counting". We openly detail the process of songwriting, arrangement, and production. This experience raised many questions on the relationship between creativity and machine, both in music analysis and generation, and on the role AI could play to assist a composer in their work. We experimented with *AI as automation*, mechanizing some parts of the composition, and especially *AI as suggestion* to foster the composer's creativity, thanks to surprising lyrics, uncommon successions of sections and unexpected chord progressions. Working with this material was thus a stimulus for human creativity.

Keywords: Co-creativity; songwriting; artificial intelligence; structure; harmony; melody; lyrics; arrangement

1. Introduction

Music creation experiments involving an artificial system are as old as the idea of computing. As early as 1843, Ada Lovelace, in her enlightening comments on the Babbage machine, imagines that the computer may *generate* music, provided that music can be *modeled*:

> *It might act upon other things besides number, were objects found whose mutual fundamental relations could be expressed by those of the abstract science of operations, and which should be also susceptible of adaptations to the action of the operating notation and mechanism of the engine... Supposing, for instance, that the fundamental relations of pitched sounds in the science of harmony and of musical composition were susceptible of such expression and adaptations, the engine might compose elaborate and scientific pieces of music of any degree of complexity or extent.*

As soon as computers became real, audio experiments began. In the late 1940's Alan Turing described how to use the *hoot* of the Mark II to produce sounds (Copeland and Long, 2017). In 1955-56, several projects generated *notated musical content* with computers, including a program to apply the compositional combinatorial rules that make the 18th-century dice game attributed to Mozart, "Musikalisches Würfenspiel" (software written by Caplin and Prinz), the song "Push Button Bertha" by Klein and Bolith written through Monte Carlo sampling from rules (Ariza, 2011), and the famous "Illiac Suite" generated with Markov chains by Hiller Jr and Isaacson (1957). These early attempts at music generation made the news already back then: both "Push Button Bertha" and the "Illiac Suite" were featured in newspaper and television programs.[1]

Throughout the years, dedicated events or societies have been encouraging the creation of pieces of music involving computers, as for example The Computer Arts Society established in 1968 or, more recently, the 2017 CrowdAI music generation challenge.[2] This is a consequence of, and a motivation for, the large body of works with an increasing variety of approaches that appeared since those pioneering efforts. For example, Fernández and Vico (2013) comprehensively survey many papers on algorithmic composition, and identify six different broad approaches: Grammars; Symbolic,

* Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL, F-59000 Lille, FR

† Normandie Univ., UNIROUEN, LITIS, F-76000 Rouen, FR

‡ Université de Picardie Jules-Verne, MIS, F-80000 Amiens, FR

Corresponding author: Mathieu Giraud (mathieu.giraud@univ-lille.fr)

Knowledge-Based Systems; Markov Chains; Artificial Neural Networks; Evolutionary and Other Population-Based Methods; and Self-Similarity and Cellular Automata. While many of these automatic composition systems are either based on rules or on machine learning techniques (Herremans et al., 2017), music generation today seems to be predominantly approached through the use of deep learning methods (Briot et al., 2019; Yang et al., 2017; Dong et al., 2018; Ji et al., 2020).

In parallel with the progress of algorithms for music generation, there has been a growing interest in machine creativity (Miller, 2020) and the difficulty of its evaluation (Jordanous, 2017). As an alternative to fully automated music generation, which transfers the whole creative task to the machine, co-creativity implies that an algorithm is rather used as a *tool* by a composer (Esling and Devis, 2020) – and this requires some steerability of the AI tools (Louie et al., 2020). As an example, co-improvisation systems (Assayag et al., 2010; Gifford et al., 2018) such as ImproteK (Nika et al., 2017) are usually based on a real-time interaction between the human musicians and the machine. Each performer (human or machine) listens to the music produced by the other and responds appropriately, bringing on the musical discourse in a novel way each time. Co-creativity can be used also in the context of a compositional process. The AI system generates a set of musical fragments, one of which will eventually be selected and possibly re-shaped by the composer to meet a specific musical need. This is the technique used by Ghisi (2017), who chooses among fragments generated by the Long Short-Term Memory neural network described in SampleRNN (Mehri et al., 2016).

Research on human-computer interaction promoting creativity notably includes a study from Lubart (2005). Lubart's classification contains the *computer as a colleague*, where the human and the machine complement each other during the creative process. One of the tactics presented by Lubart is when the machine efficiently computes multiple random searches in a possibly constrained space. Selecting the best output and transforming it into a plausible creative production is then likely to be done better by humans. An example of such human-computer collaboration is the composition of the album *Let's Have Another Gan Ainm* (Sturm and Ben-Tal, 2018) for which the musician Daren Banarsë selected melodies generated with the folk-rnn system and then modified them (Ben-Tal et al., 2020). The idea of *co-creative colleague*, as well as the notion of a *creativity support tool*, are also described as roles for AI in the creative process by Kantosalo and Jordanous (2020).

However here computers and humans do not stand on equal footing. We call here *"AI as suggestion"* the situation when such a computer colleague *suggests ideas or solutions* for a set of compositional sub-tasks, and then the composer would systematically have the final decision. In the co-creative experiment described in this paper, we adopt this approach for the melody, but also extend it to the generation of additional musical layers, namely chord sequences, lyrics, and global structure.

Contents

This manuscript describes the journey of an MIR research team experimenting with human/AI co-creation to write a song for the AI Song Contest 2020. This contest gathered 14 international teams that competed to produce a song in the style of Eurovision pop through the use of artificial intelligence (Huang et al., 2020).

Our guideline was thus to experiment with "AI as suggestion" for as many layers as possible, not only for melody, but also for chords, lyrics, and even structure. As far as we know, this kind of approach is rather unique.

Given the singularity of the AI Song Contest challenge, and in particular its aesthetic dimension, the nature of this manuscript might slightly differ from usual studies in music information retrieval or music generation. The competition schedule was tight and the team knew from the beginning that completing a song would require to span a variety of music generation tasks that did not necessarily fit with the team's expertise. Some methods have therefore been elaborated, re-implemented and customized to our needs, possibly with simplifications or hacks and without exploring all the alternatives in the literature. The unconditional need for a final song having a minimum perceptible level of *Eurovisionness* pushed the team to focus more on the co-creative process than on an evaluation of models.

As seen above, there has been quite some media buzz around "AI-generated songs", especially in recent years. However, it is often difficult to distinguish what is AI-generated from what is achieved through human intervention. In this paper, we decided to be fully transparent about our process; this includes pointing out all the hacks that we implemented since they are, in our opinion, essential to the co-creative experience in the spirit of using the *computer as a colleague*. The acts of *modeling* and *generation* are here followed by acts of *selection* and *composition*. The choice of this approach enables us to promote the ability of the machine to push composers beyond the usual boundaries of their imagination and bring novelty to their compositions.

The paper is structured as follows. Section 2 presents the approach and the timeline. Sections 3 and 4 present the co-creative approach used first in *song composition*, then in *song arrangement and production*. Sections 5 and 6 finally sum up what we learned from the experience and our thoughts on co-creativity.

2. Team, Approach, and Process

The team is composed of five MIR researchers working on music modeling (especially music analysis), with a focus on structure, harmony, and texture. Previous work of the team includes *form analysis* of fugues (Giraud et al., 2015) and sonata forms (Allegraud et al., 2019). Most members of the team have pursued academic studies in classical music and some have experience also in folk, choral, or electronic music. All of the members sing or play one or more musical instruments, including drums, accordion, guitar, flute, recorder, violin, piano, and other keyboards. Their musical experience is what allowed them to choose a co-creative approach because they could act

both as scientists and as amateur musicians according to the needs. The team would like to thank the guest singer, Niam, who is a student at the same university.

While some musical decisions were deliberately conceded to the computer, the team never intended to produce a fully 100% "just-push-a-button" submission. They prefer that artificial intelligence assists the composers, instead of substituting them. The goal was therefore to experiment with the concept of *co-creativity* between the AI and the human, which will be further discussed in Section 5. The team saw two ways in which AI could assist a composer:

- *AI as automation.* AI could liberate the composer from some compositional sub-tasks and decisions, typically to allow him to focus on some, possibly more creative, other ones.
- *AI as suggestion.* The use of AI could be limited to *suggesting solutions* for a set of compositional sub-tasks, whereas the composer would systematically have the final decision.

While the first approach falls within the current trend of considering AI as a substitute for the human, like in image classification or autonomous driving, the second one seems more specific to the artistic field. In the classification described by Lubart (2005), *AI as suggestion* matches the idea of a *creative act through integrated human-computer cooperation during idea production*, which includes the need of a final human selection and honing among computer outputs. Following the categorization by Kantosalo and Jordanous (2020), we use AI as a creativity support tool to *support* and *enable* the composer, but also and especially as a co-creative colleague, delegating most of the *generation* to it.

These two distinct uses of AI for music composition also differ on the question of evaluation. When AI is used to automate a process, the algorithm is expected to be *performant*. Machine learning algorithms typically come with a range of metrics that enable the evaluation of the performance thanks to ground truth values of a test set. However, evaluation is more delicate when the algorithm is expected to be *creative* as it is the case when outputting suggestions that are evaluated with aesthetic, and therefore subjective, criteria.

On the subject of what was expected from the teams, the competition rules originally stated:

> The AI Panel will judge the songs based on different levels. How was the provided **dataset** used? Has the song an interesting **structure**? To what extent have the **melody**, **harmony**, **lyrics**, and **audio rendering** been generated? The more elements are created with AI, the more points you will earn from the AI panel. Human interventions are allowed but this will cost you points from the AI-panel.

The original plan was to consider the 5 compositional layers mentioned in the original rules (structure, melody, harmony, lyrics, and audio rendering) and to generate each

of them independently with AI. Then, the team would manually combine – or better, *compose*, in the original Latin sense of "to put together" – the different layers. But the final rules were different:

> Effective/creative use of AI: 6 points
> Expansion of creativity: 2 points
> Discussion on co-creativity: 2 points
> Diversity and collaboration: 2 points

This evolution of the rules affected our approach to co-creativity which gradually mutated during the contest (and, honestly, even after, while writing this article). It accentuated our desire to investigate and understand how AI may suggest unexpected objects rather than automate some processes.

Altogether, to measure the human participation in the process, we carefully kept track of all human interventions for the song composition and report them in the next section using the tag[human]. This tag occurs more than 25 times, which emphasizes this co-creative approach.

3. Song Composition

Generating structured music, especially involving long-term correlations between elements, is a key challenge in music generation (Herremans and Chew, 2017; Medeot et al., 2018; Zhou et al., 2019; Dhariwal et al., 2020).

The team decided to tackle this problem by generating and selecting a song structure as the first step of the compositional process (Section 3.1), then conditioning the generation of the remaining four layers on this reference structure. We thus refer to this method as a *structure-based* approach. The structure obtained in the first step consists of a sequence of section labels such as "chorus" or "verse". In the second step, several chord sequences were generated and selected for each section of the structure (Section 3.2). Lyrics, melodies and hooks were finally generated and selected (Sections 3.3 and 3.4). For each of these elements, the co-creative approach consisted of two successive steps: *model/generate* and *select/compose*.

- *Model preparation/content generation* (about one month long). Structure, harmony, and hook melody were *modeled*, based on corpus data and methods coming either from the team's research background or from the literature. These models, as well as an available pre-trained model for lyrics, were used to *generate* new data. Models were iteratively tested and refined to improve the quality of the generation based on subjective evaluations of the outputs[human]. However, no decision on the actual music content was taken during this stage.
- Actual *song composition* (mainly on day D-21). The team gathered one morning, and it is perhaps during that meeting that the co-creative approach was most obviously followed. The team decided to *filter out*[human] some of the AI-generated material but also to give the final word to chance and to *select* among the remaining alternatives by *rolling the dice*.

Machine learning involves the estimation of probability distributions from training data. Using probabilities or randomness seems to fall under the jurisdiction of *AI as automation*, notably when sampling from these distributions. However, the dice rolls in our song composition were also a stimulus to the creativity for the subsequent steps. Musicians often have to select between original chord patterns, lyrics, and melodies that they can fluently create in their mind. Here these arbitrary constraints that we put on ourselves made the whole endeavor at least as stimulating as a completely unbounded creative process. This is a known phenomenon and lots of artists have used it, including in music (Eno and Schmidt, 1975). Therefore, we see this whole filtering-and-selection process as a typical task of co-creativity and a textbook example of *AI as suggestion.*

More generally, in the review paper of the contest, Huang et al. (2020) reveal that every team used such an approach, therein defined as *AI-generated, human curated content*, on at least one of the aspects of the final composition (Huang et al., 2020, **Figure 1**). In the following, we dedicate a section to each of the four layers that pertain to symbolic music composition: structure, chords, lyrics, and melody.

3.1 Structure
Dataset: SALAMI (https://github.com/DDMAL/salami-data-public). 2000+ structures, including 400+ pop titles (Smith et al., 2011); and the 6 labels of the Eurovision dataset, which was provided by the organizers of the
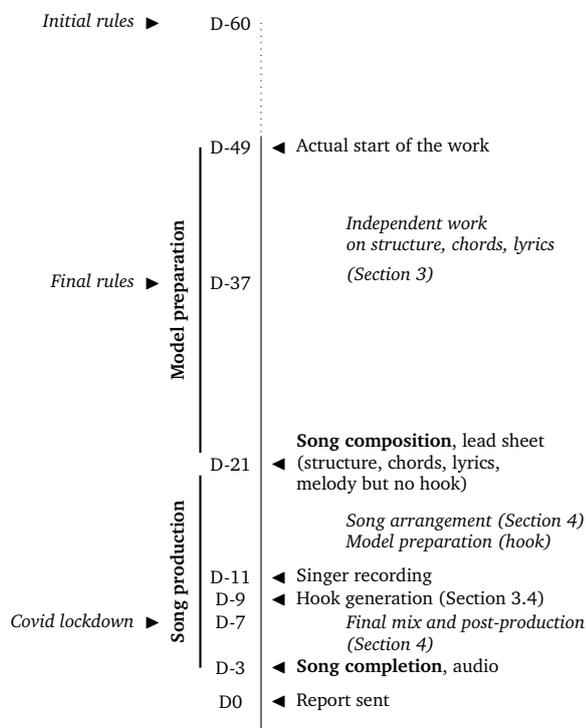

Figure 1: The preparation of the composition models took 28 days (D-49 to D-22), before the actual day (D-21) of the song composition. The song arrangement and production was then done in about 18 days, until D-3.

contest: `intro`, `verse`, `bridge`, `pre-chorus`, `chorus`, and `hook/instrumental`.

The studies involving music structure cited previously do not explicitly build a structure, but rather include structure in the learning and generation of music. Instead, the team decided to generate a structure template separately with a dedicated model.

Model, Generate. SALAMI provides structures for a large set of songs with a diverse vocabulary. The dataset was simplified by ignoring section durations and merging consecutive identical labels. The SALAMI labels were mapped[human] to the most frequent labels of the Eurovision dataset. For example, `interlude`, `transition`, `bridge`, but also all `{pre,post}-{chorus,verse}` SALAMI labels were mapped to Eurovision label `bridge`. The training and the generation were performed by a random walk on a first-order Markov model learning the succession of sections, starting and ending in special `start` and `end` states (not shown below). The model was constrained[human] to generate structures containing between 4 and 9 sections, and with at least 2 sections repeated at least 2 times. There are more than 100,000 different structures that could be generated with the given Markov model and such constraints. The model generated the following 20 structures:

S1   [intro, verse, bridge, chorus, bridge, chorus, hook]
     [verse, chorus, verse, chorus, hook, verse, chorus, hook]
S2   [intro, verse, chorus, verse, chorus, hook]
S3   [intro, verse, chorus, bridge, verse, chorus, bridge, chorus, hook]
     [intro, bridge, chorus, bridge, chorus, hook]
S4   [intro, verse, chorus, verse, chorus, hook, chorus, hook]
S5   [intro, verse, chorus, verse, bridge, hook, verse, chorus, hook]
S6   [intro, verse, chorus, hook, chorus, hook]
     [intro, bridge, chorus, bridge, chorus, hook]
S7   [intro, verse, chorus, bridge, verse, chorus]
     [intro, bridge, verse, bridge, verse, chorus, verse, chorus, hook]
     [intro, verse, bridge, verse, hook, verse, chorus, hook]
S8   [intro, chorus, verse, bridge, verse, chorus, bridge, chorus, hook]
S9   [intro, verse, chorus, verse, chorus, verse, bridge, verse, chorus]
     [intro, verse, bridge, verse, bridge, verse, chorus, hook]
S10 [intro, chorus, verse, chorus, verse, chorus, hook]
S11 [intro, verse, chorus, verse, bridge, verse, chorus, hook]
     [intro, bridge, chorus, bridge, chorus, chorus]
     [intro, verse, bridge, chorus, verse, bridge, chorus, hook]
     [applause, intro, bridge, verse, bridge, verse, chorus, hook, hook]

Note that some rare labels in SALAMI, such as "applause", appear in the outputs because the team forgot to map them to Eurovision labels.

Select, Compose (Day D-21). Following the *AI as suggestion* paradigm, the team discarded[human] some of the generated structures and kept only 11 candidates, the ones marked above from S1 to S11. A dice roll selected the structure S8, **[intro, chorus, verse, bridge, verse, chorus, bridge, chorus, hook]**. This structure didn't seem particularly natural to the team, especially because of the bridge appearing once between two verses and once between two choruses, and we wondered whether we should have removed this structure from the selected list. By eventually deciding to keep S8, the team experienced a particular creative constraint that would not have appeared if the selected structure was more conventional (e.g. S4).

Since the team discarded section durations from the input SALAMI data, they decided[human]– after the lyrics generation/composition, see below – to duplicate some of the sections, namely the first verse (Verse 1a/1b) and

the last chorus (Chorus 3a/3b). Hence, the final structure is **[Intro, Chorus 1, Verse 1a, Verse 1b, Bridge 1, Verse 2, Chorus 2, Bridge 2, Chorus 3a, Chorus 3b, Outro]** (**Figure 4**).

3.2 Chords

Dataset: Eurovision MIDI (as provided by the organizers), 200+ songs.

There is a fair number of studies on chord sequence generation, including recent works with generative grammars or deep learning (Conklin et al., 2018; Huang et al., 2016; Paiement et al., 2005; Rohrmeier, 2011).

Model, Generate. The team decided to build a model whose training would be limited to the chord sequences of the 200 songs of the Eurovision dataset. Performing effective learning on such a small dataset was a major challenge of this task. We simplified the problem by taking into account the standard sections in pop song structures, as used above, and we decided to generate one short chord sequence per section instead of a very long one covering the entire song.

As the dataset was relatively small, every chord sequence was transposed to C major or A minor after key detection with the Krumhansl-Schmuckler algorithm (Krumhansl, 1990) implemented in music21 (Cuthbert and Ariza, 2010). Chords were then encoded with two one-hot encoded vectors, the pitch class of the *chord root* and the *quality* of the chord, estimated with the *pitchedCommonName* method of music21 from the chord parts provided in the dataset. Eleven different chord qualities were considered, the ten most common in the dataset (`major triad,` `minor triad,` `minor seventh chord,` `dominant seventh chord,` etc.) as well as a catch-all value `other chord`. A small neural network, made of a single LSTM layer with 40 hidden units followed by a dense output layer, was trained on this dataset and used to generate chord sequences.

The model generated the following sequences, with sharps sometimes rewritten as flats for better understanding (nc means "no chord" and question mark stands for "other chord"):

Intro (2/10)
Cm A9 F F? GØ7 Fm Em7 G D7 (…)
I1   **C Gm7 A Cm7 C A Cm Bm7 C9 C E9 (…)**
GMaj7 EMaj7 A♯m F♯Maj7 Bh7
Cm Fm G9 FMaj7 Fm7 Gm E C D A (…)
G♯m F7 Bm7 Gm9 Am9 E7 Dnc (…)
E? D♯? Gm7 C A G♯ F Bm (…)
G FMaj7 Am F♯ A? D? F9 Cm7 Fm Cm9 (…)
G2 G♯7 D7 Em Cm G Am A♯
Am
I2   **C9 Cm7 B♭m7 G? Bm Dm7 Fnc AMaj7 A7 (…)**

Verse (3/10)
V1   **C C F F? Cnc**
F E9 Gm
Am9 F? Eh7 G♯m7 A♯m7 A♯m C Gnc Gnc CMaj7
?
V2   **C F Am Em A♭ C Anc Am Cm C? Dm Dnc**
G Dm7 G7 Cm Dm Bm A F♯m
V3   **C D F♯Maj7 Gnc**
A Em G G♯ Eh7 D♯m9 G Gh7 Ah7 Gm (…)
Am F A♯7 Em Gm9 Dm Dm G♯ Cm
C Bnc C♯m A9 F9 Dh7 Cm F9 A♯m9 E (…)
    (…)

Some of these chord sequences are barely tonal, but some others seemed reasonably common to the team. Although the model and the generation process could be improved in many ways, the team estimated the outputs exploitable enough to move on to the next stage.

Select, Compose (Day D-21). Following the same method employed for the structure, the team selected[human] a few chord progressions (13 in total, from a pool of 10 sequences for each of the five sections) and rolled the dice to get:

Intro: C Gm7 A Cm7 C A
Chorus: Am Em Gm Cm B♭ m Cm D Fnc
Verse: C D F♯Maj7 Gnc
Bridge: Em D A C?
Hook: C F Dm Fm

Human adjustments. The team eventually decided[human] to keep for the Intro the same chord sequence as for the Hook, with a C pedal. Since the passage from F♯Maj7 to G felt rather uncommon, it was decided[human] to swap the sequences of the Bridge and the Verse in order to limit its occurrence. The team still decided to keep the "unexpected" chords that the AI suggested, such as B♭m in the chorus and F♯Maj7 in the bridge.

The harmonic rhythm (*i.e.,* chord durations) was not generated together with the chords themselves. Different choices were made depending on the sections, but always with a regular rhythm, such as "one chord every measure", inside each section. An exception was made[human] for the chorus, in which we gathered 4 chords in a single measure to facilitate the inclusion of the unexpected chord B♭ m chord in a "passing chord" style. This choice also had the consequence to close the chorus[human] on the bright D major chord. The final chord sequences, with the harmonic rhythm, are thus:

Hook	C	F/C	Dm/C		Fm/C
Chorus	Am	Em	Gm Cm/A B♭m Cm	D	
Verse	Em	–	D	–	
	A	–	C	–	
Bridge	C	D	F♯Maj7	G	

Except for the bridge, the team decided[human] to loop twice over the chord sequence in each section, and to double the length of the verse to leave place for the lyrics (see next section). The piano rendering of the chords was a voicing[human] of these chords, which tried to put in relevance the B♭m and F♯Maj7 chords (see Section 4).

A spurious B♮. The team realized at D-9, about one week before submission, that they made an unintended variation[human] in the piano part, playing a G major (and keeping this B♮ a few beats) instead of the Gm generated by the model at the second last measure of every chorus (labeled as G* on **Figure 4**). Note that they worked on the song for two weeks before realizing such a "mistake"! Even if the last chords of the chorus (B♭m Cm D) may imply a B♭ note and therefore a Gm harmony, the chorus,

that can be heard in A aeolian or A dorian, calls here for a G major harmony. Being exactly in between these two progressions, this chord can fit in equally well when played in both modes. Musicians sometimes auto-correct things without realizing it (Sloboda, 1984). When the team saw this variation, the vocals were already recorded and the song was almost ready. The team decided[human] to keep this unexpected artifact as a manifestation of the co-creative approach.

3.3 Lyrics

Dataset: Eurovision Lyrics (as provided by the organizers), 200+ songs.

Can the spirit of the lyrics of Eurovision songs be captured by an AI model? As the lyrics generation had been done independently from the musical composition, our first experiments yielded texts without "musicality", and notably with an irregular number of syllables in different verses. The most common text models may be more targeted at semantics than metrics: That is why some music generation studies target lyrics generation with controlled rhythm and meter (Barbieri et al., 2012).

Model, Generate. The team then tested an extreme position: can Eurovision songs convey insightful messages with only two words? To answer this question, a list of all pairs of words (bi-grams) in the Eurovision dataset was produced, focusing on nominal groups and complete sentences (see **Figure 2**). There was no constraint on the number of syllables, but it turned out that *all* words appearing in the 100 most frequent bi-grams have a single syllable. The team kept almost[human] all of them, and, with a small reordering[human], obtained a seed (top of **Figure 3**) that was then used as an input for the GPT-2 model (Radford et al., 2019) to generate longer lyrics. Such a generation based on bi-gram statistics prevents the risk of plagiarism. Who could say that the team copied verses such as *my heart* when they are already used by hundreds of songs? However, this conformism can arguably contradict

in the	192	*my heart*	124
and i	129	i'm	124
my heart	124	i can	118
i'm	124	we are	116
i can	118	i am	84
we are	116	i know	82
like a	100	you are	82
oh oh	99	*my love*	58
don t	94	my life	55
to the	93	*the world*	53
to me	89	*you know*	51
i am	84	*your eyes*	50
on the	84	*your heart*	42
in my	83	*thank you*	42
i know	82	the way	39
you are	82	*the sky*	37
of the	82	my mind	37
if you	77	*the sun*	35

Figure 2: Most frequent bi-grams in the Eurovision lyrics dataset along with the number of occurrences each. (Left) All pairs of words; (Right) Nominal groups, or complete sentences. Italic words were selected to make the seed.

the quest for creativity. This type of tension between conformism and creativity happened to appear a number of times along the whole co-creativity experiment.

Select, Compose (Day D-21). The first generation gave a text that the team split[human] into blocks for each section (**Figure 3**). The team liked such repetitiveness, but also the call/response between "I *stop* counting" and "I *keep* counting" that the team assigned[human] to the chorus (and completed[human], see **Figure 4**, to also repeat "I keep counting" twice). Repetitiveness is typically undesirable in prose, but actually quite musical and appropriate to the lyrics of a song (or even a poem).

3.4 Melody

There are many recent approaches to melody generation, possibly constrained by underlying chord sequences or pre-existing lyrics. Most of them use machine learning (Pachet and Roy, 2011; Shin et al., 2017; Tardon-Garcia et al., 2019; Yang et al., 2017; Zhu et al., 2018). However, having largely used artificial processes to produce the previous layers, the team decided to favor a human approach on this particular task.

During the meeting on composition day (D-21), after the lyrics were fixed (see above), one person played the chords on the piano and the rest of the team gathered around a table and hummed along[human] until a simple and catchy melody appeared. For the verses, no melody was composed (see Section 4) other than choosing the rhythm of the bi-grams as two eighth-notes on the first beat of every measure. The team further decided at D-21 to use AI to generate a melody for the *Hook* instrumental section.

Hook Generation. Dataset: around 10,000 melodies of common practice period coming from *A Dictionary of Musical Themes* (Barlow and Morgenstern, 1948), previously available at http://www.multimedialibrary.com/barlow.

Since the goal was to produce an instrumental track, the team decided[human] to use the instrumental musical themes in the database by Barlow and Morgenstern. Even if they are probably not coherent with the Eurovision style, they offer a nice change of perspective from the dataset that was used so far and also, conveniently, a vastly larger set of files to train on.

Model, Generate. The tonality of the themes was estimated with music21. All themes in a minor key were

Verse 1a and 1b
my heart, my love
the world, you know,
your eyes, your heart,
thank you, my soul

Verse 2
One day, the sun
the sky, no one
I know, at least you come,
Each day when we go
on the war
between mine and yours,

Chorus 1
I stop counting
I stop counting,
It will never end

Chorus 2
I keep counting
I count down the years

Bridge
Oh, it's ending, it's ending
It's ending, it's ending

Figure 3: Lyrics, in the order they were generated by GPT-2. The seed is in italics.

then discarded[human] and everything else was transposed to C major. Statistics of note durations and intervals between notes and tonic were computed. The team then used these distributions to sample sequences of notes with a total duration of 8 measures. As Baroque, Classical and Romantic themes in the selected dataset generally have more notes than pop songs, the team multiplied[human] all note durations by 2 and fixed the minimal generated duration to a quaver. To limit excessive melodic gaps, they forced[human] the generated notes to the range between F♯ below and the F above middle C.

(Contd.)

Figure 4: The lead sheet of "I keep counting" resulting from the process described in Section 3. The G* chord in the chorus is discussed at the end of Section 3.2. The slashed bars in the verses indicate where the melody was not fixed before the recording session (see Sections 3.4 and 4).

This model disregards all internal music structure as well as the voicing, but produces some plausible melodies. With such a sampling, it is expected that diatonic tones, and especially the notes of the tonic and dominant triads (C/E/G/B/D) play a more significant role. This follows the known patterns of pitch profiles (Krumhansl and Kessler,

1982; Temperley, 1999), However, the generated melodies included other notes as well.

Select, Compose (Day D-9). Out of 20 generated sequences (available in supplementary material), the team selected[human] the one most befitting the existing chord progression according to their musical taste. The team decided [human] to consider the first 4 measures of this generation as two phrases, and to loop twice over them (**Figure 4**). A further dice roll would not have fostered the co-creativity anymore as the song was, at this point, almost finalized.

4. Music Arrangement and Production[human]

Music arrangement, orchestration, and mixing form a crucial part in the composition of a piece of popular music. Automatic mixing and production form a growing research area (Deruty, 2016; Man et al., 2017; Birtchnell, 2018). Although the team decided not to use any AI method for this layer, human interventions were intended to be *as discreet as possible* to avoid the AI-generated content being pushed into the background. This decision was mostly due to the initial rules of the contest, with the objective to be able to identify all the contributions during the creative process. The music production could have blurred previous choices with too prominent effects. Thus the team did not trust an automated tool, but human choices at this stage could have had the same effect. After such a first contest, we would probably today make other choices and seek a co-creative blend between AI and human content.

The team thus used its few skills in music production and favored whenever possible default settings to limit human decisions. Although not directly related to AI, this approach also follows the *automation* dimension of our interpretation of the AI song challenge as discussed in Section 2. As an exception, the team decided to give the lead vocal part to a human singer.

Piano/pad, bass and strings. The lead-sheet was notated in MuseScore[3] with some instrumental tracks, exported as a MIDI file, and then opened within the Digital Audio Workstation Logic Pro X (LPX). The notes played in the piano track in every section result in a human-made voicing of the generated chord sequences (Section 3.2) with occasional additional non-chord notes. The rhythm of this track is relatively deterministic. In most of the song, the piano notes are played just on the onset of every chord. On Verse 2 and Chorus 3, they are played on every eighth-note. The piano track was rendered with the *Yamaha Grand Piano* virtual instruments (VSTi) of LPX. At the end of the song (Bridge 2 and Chorus 3), this piano track is duplicated with an arpeggiator MIDI effect and accompanied by an additional pad track, rendered with the *Pad* VSTi, which plays the same notes.

A bass line and string line were also composed by humans. The bass plays the root note of the chords − except on the measure with unexpected chords, see section 3.2 − and adds some straight fills between chords and between sections, that evolve along the sections of the song. The string line also underlines some sections and transitions, playing mostly chord tones in half notes,

sometimes again with short fills. The LPX *Subby Bass* VSTi was used to render the bass track and the *Modern Strings* VSTi to render the string track.

Hook. The *African Kalimba* VSTi was used to render the intro/outro hook. As mentioned earlier, it was generated almost at the end of the work (D-9) and at this time, the team had an increasingly personal and precise idea of the targeted final song, selecting a sound that fits to the song. The time was also running short, and the first nationwide Covid-19 lockdown surprised everyone and disorganized the work on the song a bit.

Tempo, percussion, and drums. The tempo was chosen at 128 bpm, which is in a standard range for pop and dance music. A percussion track (Intro, Chorus 1, Verse 1a, Verse 1b) and a drum track (Verse 1b, Bridge 1, Verse 2, Bridge 2, Chorus 3, and Outro) were designed with the assistant *Drummer* tool of LPX which provides predefined loops. The team manually selected loops in the *Darcy − Retro* (for drums) and *Quincy − Studio* (for percussion) categories. A few high-level settings (complexity, fills, swing, etc.) were manually added.

Vocals. The female singer, Niam, is the only live (human) musician performing the song. The vocals have been recorded with a microphone Warm Audio WA-47jr in a room dedicated to audio research and recording from the *Science and Culture du Visuel* lab in *La Plaine Image* (Tourcoing). During the recording session, the singer semi-improvised the verse melody, freely choosing, on each measure, chord notes for the two eighth-notes, and improvising more for the second half of Verse 2. The team chose to both alternate and stack that voice with two synthetic voices (plugin *Emvoice* and the LPX vocoder *Evoc*), as well, on the Chorus 3a/3b, with a pitch-shifted voice of herself.

Mixing and mastering. Playing a background role from our point of view, and being human-composed, the team tried to keep the piano/pad, string and bass tracks as discreet as possible. The final mix and the mastering was done by a professional sound engineer.

Discussion. Even if the LPX Drummer tool, the virtual instruments, and the voice synthesis effects do not involve any AI process (at least in the way we used them), the team thought it could be interesting in the context of this competition to take advantage of these tools that emulate human performance.

To underline the structure that was generated by AI, we collectively decided how to introduce the additional tracks and voices in the various sections in order to bring global contrasts and tension progression through the song. This objective was also targeted by the final mixing and mastering.

Altogether, we recognize that the biggest human intervention here is the *arrangement plan*. Even if the team was almost the only one in the contest to generate structure from a (very simple) AI method (see Section 3.1, and Huang et al. (2020, **Figure 1**)), many decisions on the arrangement were human-made, specifically which instrumentss and accompaniment patterns enter and when. Although new services helping composers with these tasks are mainly black boxes, automatic arrangement

and orchestration is an active research field (Abreu et al., 2016; Crestel and Esling, 2017; Tsushima et al., 2018) that will bring new possibilities in the next years.

5. Discussion

Evaluating computational music creativity is difficult (Jordanous, 2012, 2017; Agres et al., 2016). Due to the tight schedule of the AI Song Contest 2020, the team did not have the time to investigate this subject thoroughly and decided to collectively compose something according to "what they like", also trying to stick with their idea of what could be a "Eurovision style". This decision introduced some risks for two different reasons. First of all, each member of the team had a different musical background and different biases that influence their tastes; as always, when different opinions meet in creative matters, the result is not necessarily equal to the sum of its parts but could be vastly superior or vastly inferior. Perhaps more importantly, however, nobody in the team was an expert in the pop style that characterises Eurovision songs.

Despite these difficulties, we were reasonably happy with the result. We thought that casual listeners could not even notice that the song was composed thanks to the assistance of AI (although it could be debated whether the dissimulation of the intervention of AI should be considered as a good point or not). Certainly, there were many strange things in the resulting song, like the slightly ungrammatical lyrics, the unusual structure, or a couple of notes that feel out of place, but they concur to produce a song that could blend in with some of the Eurovision proposals: the song was appreciated by both the public and the technical jury and it reached 4th place.

What would have been the song with other co-creative approaches is a fascinating question, but hard to answer. The next paragraphs discuss the role of constraints in artistic endeavors, our self-identification in this contest as scientists and amateur musicians, as well as questions on the intellectual property of such a song.

5.1 Creative Freedom, Artificial Constraints, and Co-Creativity

During the creative process, we often found ourselves faced with the same very old question of *creative freedom* vs *constraints*. Constraints, in some sense unintuitively, can often help an artist to achieve a more creative result, both in the case that the artist is human (Eno and Schmidt, 1975) and machine (McKeown and Jordanous, 2018). In this latter case, the easiest way to impose constraints is to introduce rules that must be respected at all times, such as *never generate a note that does not belong to the reference scale*. Doing so guarantees that there is a certain coherence between the generated piece of music and the set of all music experienced by the listeners, therefore establishing some basic facts by which the produced material can be perceived as *creative* instead of simply *random* and *rambling*. Basically, one prevents the machine from ever making *mistakes* (even though the very concept of mistake is quite ambiguous in creative endeavours). Those kinds of rules, however, also tend to produce results that quickly become quite predictable, therefore less creative.

Reybrouck (2006) studies creativity from the point of view of cybernetics theory. Creativity – and constraints – can be found in the way the musician, as a "device", processes information, but also, in the case of a "structurally adaptive device", in the very way of how she processes input or output. In our case, the expectations of the team on the different layers certainly played a role in the very way of how they even heard the suggestions from the machine. Indeed, the use of AI can be regarded, especially in the co-creative approach that we used, as a type of constraint that maximises the creativity: *AI generates a set of candidate musical objects* that limits the choice of the composer. During the entire process, the team tried to adhere as much as possible to these constraints arising from *AI as suggestion*. Several times, the team members had to unexpectedly invoke their creative skills to solve puzzles raised by such AI outputs. This phenomenon allowed the song to reach a final state that would certainly not have been possible without the intervention of AI. Finally, Todd Lubart (2005) describes AI creativity programs that *fail* at preventing a human intervention at some point as *"successful human-computer interactions to facilitate creativity"*. The human intervention that characterizes the *AI as suggestion* principle is here a deliberate choice and could not be qualified as a *fail*, but it still enables to reinforce the human-computer interaction.

5.2 Scientists and/or Artists

Although there were continuous interactions between the members of the team, the *Model preparation* tasks for the different layers between D-49 and D-22 were split among different people. This allowed for enough space for each member to contribute meaningfully with relative liberty of action. On the other hand, during the *song composition* on day D-21 and the *song production* after that, the team worked together. The members of the team often disagreed: some decisions were taken by consensus and some others not. In particular, some members advocated for hearing clearly what was AI (and, as explained on Section 4, that was the initial collective decision), others favored the aesthetics of the song.

At some point, the team felt that a clear artistic lead was missing: the team identifies itself as MIR scientists and amateur musicians. If the co-creative project had been led by a professional artist if would most certainly have evolved in a completely different direction – and the team will seek to collaborate with artists in future participation in such contests.

5.3 Data Availability

Who owns "I Keep Counting"? Intellectual property is challenging when using AI methods (McCormack et al., 2019). Most of the time, direct plagiarism is prohibited – as the rules of the AI Song Contest explicitly stated – but when does plagiarism start? Deep learning often makes it difficult to know what influenced the output. Here, as we used generally simple techniques, we are somewhat proud to positively guarantee the provenance of some elements: For example, regarding the lyrics, in the verse, *My heart, my love, the world, you know*, the bi-grams we used come

from 50+ Eurovision songs but no existing song contains all of them. Our common belief is that, if composing is selecting (with bad or good taste), the team as humans has the ownership, since we took responsibility for some choices. We released, under an open-source licence (Creative Commons CC-BY-SA 4.0), at www.algomus. fr/data, the song, the lead sheet, and some of the raw outputs of the generative process that we used for the composition. The song with its annotated sections is also available from the Dezrann platform at www.dezrann.net.

6. Conclusion

An important part of the MIR community dedicates its research to how AI can *automate* the composition process. Breakthroughs in this subject improve our knowledge in machine learning but also in composition practices. To evaluate *AI as automation*, contests could be organized towards *generative systems*. For example, the participants could be asked to provide *one hundred* different songs attesting the efficiency of their automations. A few songs per team would then be randomly chosen to be sent both to a technical jury and to the public to be judged.

On the contrary, the work done on *one* song, as in the AI Song Contest, leaves a lot of space to the application of co-creative approaches that use both *AI as automation* and, especially, *AI as suggestion*. We thus described in this paper how we tackled our structure-based songwriting using a co-creative approach, following as much as possible the *AI as suggestion* paradigm on several layers of the composition of this pop song. It required the implementation, adaptation, and, often times, the hacking of established MIR and music generation techniques. The high number of human interventions that we recorded during the entire songwriting process surprised us at first. We now think that it is a clear sign of co-creativity, and we hope that we have managed to convey here this message.

Notes

Additional File

Acknowledgements

Competing Interests

References

Abreu, J., Caetano, M., and Penha, R. (2016). Computer-aided musical orchestration using an artificial immune system. In *International Conference on Evolutionary and Biologically Inspired Music, Sound, Art and Design (EvoMUSART 2016)*, pages 1–16. DOI: https://doi.org/10.1007/978-3-319-31008-4_1

Agres, K., Forth, J., and Wiggins, G. A. (2016). Evaluation of musical creativity and musical metacreation systems. *Computers in Entertainment*, 14(3): 1–33. DOI: https://doi.org/10.1145/2967506

Allegraud, P., Bigo, L., Feisthauer, L., Giraud, M., Groult, R., Leguy, E., and Levé, F. (2019). Learning sonata form structure on Mozart's string quartets. *Transactions of the International Society for Music Information Retrieval*, 2(1): 82–96. DOI: https://doi.org/10.5334/tismir.27

Ariza, C. (2011). Two pioneering projects from the early history of computer-aided algorithmic composition. *Computer Music Journal*, 35(3): 40–56. DOI: https://doi.org/10.1162/COMJ_a_00068

Assayag, G., Bloch, G., Cont, A., and Dubnov, S. (2010). Interaction with machine improvisation. In *The Structure of Style*, pages 219–245. Springer. DOI: https://doi.org/10.1007/978-3-642-12337-5_10

Barbieri, G., Pachet, F., Roy, P., and Degli Esposti, M. (2012). Markov constraints for generating lyrics with style. In *European Conference on Artificial Intelligence (ECAI 2012)*, volume 242, pages 115–120.

Barlow, H., and Morgenstern, S. (1948). *A Dictionary of Musical Themes*. Crown Publishers.

Ben-Tal, O., Harris, M. T., and Sturm, B. L. (2020). How music AI is useful: Engagements with composers, performers, and audiences. *Leonardo*, 54(5): 510–516. DOI: https://doi.org/10.1162/leon_a_01959

Birtchnell, T. (2018). Listening without ears: Artificial intelligence in audio mastering. *Big Data & Society*, 5(2): 2053951718808553. DOI: https://doi.org/10.1177/2053951718808553

Briot, J.-P., Hadjeres, G., and Pachet, F.-D. (2019). *Deep Learning Techniques for Music Generation*. Springer. arXiv:1709.01620. DOI: https://doi.org/10.1007/978-3-319-70163-9

Conklin, D., Gasser, M., and Oertl, S. (2018). Creative chord sequence generation for electronic dance music. *Applied Sciences*, 8(9): 1704. DOI: https://doi.org/10.3390/app8091704

Copeland, B. J., and Long, J. (2017). Turing and the history of computer music. In Floyd, J. and Bokulich, A., editors, *Philosophical Explorations of the Legacy of Alan Turing: Turing 100*, pages 189–218. Springer International Publishing. DOI: https://doi.org/10.1007/978-3-319-53280-6_8

Crestel, L., and Esling, P. (2017). Live Orchestral Piano, a system for real-time orchestral music generation. In *Sound and Music Computing Conference (SMC 2017)*, pages 434–442.

Cuthbert, M. S., and Ariza, C. (2010). music21: A toolkit for computer-aided musicology and symbolic music

data. In *International Society for Music Information Retrieval Conference (ISMIR* 2010), pages 637–642.

De Man, B., Reiss, J., and **Stables, R.** (2017). Ten years of automatic mixing. In *Workshop on Intelligent Music Production (WIMP 2017)*.

Deruty, E. (2016). Goal-oriented mixing. In *Workshop on Intelligent Music Production (WIMP 2016)*, volume 13.

Dhariwal, P., Jun, H., Payne, C., Kim, J. W., Radford, A., and **Sutskever, I.** (2020). Jukebox: A generative model for music. *arXiv:2005.00341*.

Dong, H.-W., Hsiao, W.-Y., Yang, L.-C., and **Yang, Y.-H.** (2018). MuseGAN: Multi-track sequential generative adversarial networks for symbolic music generation and accompaniment. In *AAAI Conference on Artificial Intelligence (AAAI 2018)*, volume 32.

Eno, B., and **Schmidt, P.** (1975). Oblique strategies. Boxed set of cards (limited edition).

Esling, P., and **Devis, N.** (2020). Creativity in the era of artificial intelligence. *arXiv:2008.05959*.

Fernández, J. D., and **Vico, F.** (2013). AI methods in algorithmic composition: A comprehensive survey. *Journal of Artificial Intelligence Research*, 48(1): 513–582. DOI: https://doi.org/10.1613/jair.3908

Ghisi, D. (2017). *Music across music: Towards a corpus-based, interactive computer-aided composition*. PhD thesis, Pierre and Marie Curie University (Paris 6).

Gifford, T., Knotts, S., McCormack, J., Kalonaris, S., Yee-King, M., and **d'Inverno, M.** (2018). Computational systems for music improvisation. *Digital Creativity*, 29(1): 19–36. DOI: https://doi.org/10.1080/14626268.2018.1426613

Giraud, M., Groult, R., Leguy, E., and **Levé, F.** (2015). Computational fugue analysis. *Computer Music Journal*, 39(2). DOI: https://doi.org/10.1162/COMJ_a_00300

Herremans, D., and **Chew, E.** (2017). Morpheus: Generating structured music with constrained patterns and tension. *IEEE Transactions on Affective Computing*, 10(4): 510–523. DOI: https://doi.org/10.1109/TAFFC.2017.2737984

Herremans, D., Chuan, C.-H., and **Chew, E.** (2017). A functional taxonomy of music generation systems. *ACM Computing Surveys*, 50(5): 1–30. DOI: https://doi.org/10.1145/3108242

Hiller Jr, L. A., and **Isaacson, L. M.** (1957). Musical composition with a high speed digital computer. In *Audio Engineering Society Convention 9*.

Huang, C.-Z. A., Duvenaud, D., and **Gajos, K. Z.** (2016). ChordRipple: Recommending chords to help novice composers go beyond the ordinary. In *International Conference on Intelligent User Interfaces (IUI 2016)*, pages 241–250. DOI: https://doi.org/10.1145/2856767.2856792

Huang, C.-Z. A., Koops, H. V., Newton-Rex, E., Dinculescu, M., and **Cai, C. J.** (2020). AI Song Contest: Human-AI co-creation in songwriting. In *International Society for Music Information Retrieval Conference (ISMIR* 2020).

Ji, S., Luo, J., and **Yang, X.** (2020). A comprehensive survey on deep music generation: Multi-level

representations, algorithms, evaluations, and future directions. *arXiv:2011.06801*.

Jordanous, A. (2012). A standardised procedure for evaluating creative systems: Computational creativity evaluation based on what it is to be creative. *Cognitive Computation*, 4(3): 246–279. DOI: https://doi.org/10.1007/s12559-012-9156-1

Jordanous, A. (2017). Has computational creativity successfully made it "beyond the fence" in musical theatre? *Connection Science*, 29: 350–386. DOI: https://doi.org/10.1080/09540091.2017.1345857

Kantosalo, A., and **Jordanous, A.** (2020). Role-based perceptions of computer participants in human-computer co-creativity. In *AISB Symposium of Computational Creativity (CC@AISB 2020)*.

Krumhansl, C. L. (1990). *Cognitive Foundations of Musical Pitch*. Oxford University Press.

Krumhansl, C. L., and **Kessler, E. J.** (1982). Tracing the dynamic changes in perceived tonal organisation in a spatial representation of musical keys. *Psychological Review*, 89(2): 334–368. DOI: https://doi.org/10.1037/0033-295X.89.4.334

Louie, R., Coenen, A., Huang, C. Z., Terry, M., and **Cai, C. J.** (2020). Novice-AI music co-creation via AI-steering tools for deep generative models. In *Conference on Human Factors in Computing Systems (CHI 2020)*, pages 1–13. DOI: https://doi.org/10.1145/3313831.3376739

Lovelace, A. (1843). A sketch of the analytical engine, with notes by the translator. *Scientific Memoirs*, 3: 666–731.

Lubart, T. (2005). How can computers be partners in the creative process: Classification and commentary on the special issue. *International Journal of Human-Computer Studies*, 63(4–5): 365–369. DOI: https://doi.org/10.1016/j.ijhcs.2005.04.002

McCormack, J., Gifford, T., and **Hutchings, P.** (2019). Autonomy, authenticity, authorship and intention in computer generated art. In *International Conference on Computational Intelligence in Music, Sound, Art and Design (EvoMUSART 2019)*, pages 35–50. DOI: https://doi.org/10.1007/978-3-030-16667-0_3

McKeown, L., and **Jordanous, A.** (2018). An evaluation of the impact of constraints on the perceived creativity of narrative generating software. In *International Conference on Computational Creativity (ICCC* 2018).

Medeot, G., Cherla, S., Kosta, K., McVicar, M., Abdalla, S., Selvi, M., Rex, E., and **Webster, K.** (2018). StructureNet: Inducing structure in generated melodies. In *International Society for Music Information Retrieval Conference (ISMIR* 2018).

Mehri, S., Kumar, K., Gulrajani, I., Kumar, R., Jain, S., Sotelo, J., Courville, A., and **Bengio, Y.** (2016). SampleRNN: An unconditional end-to-end neural audio generation model. *arXiv:1612.07837*.

Miller, A. (2020). *The Artist in the Machine: The World of* AI-*Powered Creativity*. MIT Press. DOI: https://doi.org/10.7551/mitpress/11585.001.0001

Nika, J., Chemillier, M., and **Assayag, G.** (2017). Improtek: Introducing scenarios into human-computer music

improvisation. *Computers in Entertainment*, 14(2): 1–27. DOI: https://doi.org/10.1145/3022635

Pachet, F., and Roy, P. (2011). Markov constraints: Steerable generation of Markov sequences. *Constraints*, 16(2): 148–172. DOI: https://doi.org/10.1007/s10601-010-9101-4

Paiement, J.-F., Eck, D., and Bengio, S. (2005). A probabilistic model for chord progressions. In *International Conference on Music Information Retrieval (ISMIR 2005)*.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., and Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI Blog*, 1(8): 9.

Reybrouck, M. M. (2006). Musical creativity between symbolic modelling and perceptual constraints: The role of adaptive behaviour and epistemic autonomy. In *Musical Creativity*, pages 58–76. Psychology Press. DOI: https://doi.org/10.4324/9780203088111-13

Rohrmeier, M. (2011). Towards a generative syntax of tonal harmony. *Journal of Mathematics and Music*, 5(1): 35–53. DOI: https://doi.org/10.1080/17459737.2011.573676

Shin, A., Crestel, L., Kato, H., Saito, K., Ohnishi, K., Yamaguchi, M., Nakawaki, M., Ushiku, Y., and Harada, T. (2017). Melody generation for pop music via word representation of musical properties. *arXiv:1710.11549*.

Sloboda, J. A. (1984). Experimental studies of music reading: A review. *Music Perception*, 2(2): 222–236. DOI: https://doi.org/10.2307/40285292

Smith, J. B. L., Burgoyne, J. A., Fujinaga, I., De Roure, D., and Downie, J. S. (2011). Design and creation of a large-scale database of structural annotations. In *International Society for Music Information Retrieval Conference (ISMIR 2011)*.

Sturm, B. L., and Ben-Tal, O. (2018). Let's have another Gan Ainm: An experimental album of Irish traditional music and computer-generated tunes. Technical Report, KTH Royal Institute of Technology.

Tardon-Garcia, L. J., Barbancho-Perez, I., Barbancho-Perez, A. M., Roig, C., Tzanetakis, G. (2019). Automatic melody composition inspired by short melodies using a probabilistic model and harmonic rules. In *International Society for Music Information Retrieval Conference (ISMIR 2019)*.

Temperley, D. (1999). What's key for key? The Krumhansl-Schmuckler key-finding algorithm reconsidered. *Music Perception*, 17(1): 65–100. DOI: https://doi.org/10.2307/40285812

Tsushima, H., Nakamura, E., Itoyama, K., and Yoshii, K. (2018). Interactive arrangement of chords and melodies based on a tree-structured generative model. In *International Society for Music Information Retrieval Conference (ISMIR 2018)*.

Yang, L.-C., Chou, S.-Y., and Yang, Y.-H. (2017). MidiNet: A convolutional generative adversarial network for symbolic-domain music generation. In *International Society for Music Information Retrieval Conference (ISMIR 2017)*, pages 324–331.

Zhou, Y., Chu, W., Young, S., and Chen, X. (2019). BandNet: A neural network-based, multi-instrument Beatles-style MIDI music composition machine. In *International Society for Music Information Retrieval Conference (ISMIR 2019)*.

Zhu, H., Liu, Q., Yuan, N. J., Qin, C., Li, J., Zhang, K., Zhou, G., Wei, F., Xu, Y., and Chen, E. (2018). XiaoIce Band: A melody and arrangement generation framework for pop music. In *International Conference on Knowledge Discovery and Data Mining (KDD 2018)*, pages 2837–2846. DOI: https://doi.org/10.1145/3219819.3220105