

Article

Using Formal Grammars as Musical Genome

David D. Albarracín-Molina ^{1,*} , Alfredo Raglio ² , Francisco Rivas-Ruiz ³  and Francisco J. Vico ¹

- ¹ ETSI Informatica, Universidad de Málaga, Andalucía Tech, C/Severo Ochoa, 4, Institutos Universitarios (UMA), Campanillas, 29590 Málaga, Spain; fjvico@uma.es
² Istituti Clinici Scientifici Maugeri IRCCS, Via Salvatore Maugeri 10, 27100 Pavia, Italy; alfredo.raglio@unipv.it
³ Research Unit of the Agencia Sanitaria Costa del Sol, Hospital Costa del Sol. 5ª Planta. Autovía A7, Km 187, 29603 Marbella, Spain; frivasr@hcs.es
* Correspondence: david@geb.uma.es

Abstract: In this paper, we explore a generative music method that can compose atonal and tonal music in different styles. One of the main differences between regular engineering problems and artistic expressions is that goals and constraints are usually ill-defined in the latter case; in fact the rules here could or should be transgressed more regularly. For this reason, our approach does not use a pre-existing dataset to imitate or extract rules from. Instead, it uses formal grammars as a representation method than can retain just the basic features, common to any form of music (e.g., the appearance of rhythmic patterns, the evolution of tone or dynamics during the composition, etc.). Exploring different musical spaces is the responsibility of a program interface that translates musical specifications into the fitness function of a genetic algorithm. This function guides the evolution of those basic features enabling the emergence of novel content. In this study, we then assess the outcome of a particular music specification (guitar ballad) in a controlled real-world setup. As a result, the generated music can be considered similar to human-composed music from a perceptual perspective. This endorses our approach to tackle arts algorithmically, as it is able to produce novel content that complies with human expectations.



Citation: Albarracín-Molina, D.D.; Raglio, A.; Rivas-Ruiz, F.; Vico, F.J. Using Formal Grammars as Musical Genome. *Appl. Sci.* **2021**, *11*, 4151. <https://doi.org/10.3390/app11094151>

Academic Editor: Lamberto Tronchin

Received: 14 March 2021

Accepted: 29 April 2021

Published: 1 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: generative music; formal grammars; L-systems; evolutionary computation; genetic algorithms; music perception

1. Introduction

Creativity is a complex concept that has been forged in the modern eras when it was started to be perceived as a capability of great men, in contrast to just being productive [1]. There are a variety of definitions, but here we understand it as the ability to generate novel and valuable ideas [2]. Music composition is considered an expression of human creativity, even if composers (like artists in general) take inspiration from other sources, such as Nature sounds and -mostly- other authors. Similarly, the algorithmic composition of music usually follows an imitative approach by feeding a computer system with a large corpus of existing (human-made) scores. However, here we investigate music composition from a different perspective: as an iterative process to discover esthetically pleasing musical patterns.

1.1. Automated Composition

The digital era brought yet another step in the evolution of musical genres [3] and the development of algorithmic composition [4]. Some works at the early age were: an unpublished work by Caplin and Prinz in 1955 [5], with an implementation of Mozart's dice game and a generator of melodic lines using stochastic transitional probabilities; the Hiller and Isaacson's Illiac Suite [6], based on Markov chains and rule systems; and MUSI-COMP by Baker [7], implementing some methods used by Hillers and Xenakis's stochastic algorithms in the early 1960s, working as CAAC [7]. In 1961, a dedicated computer was able to compose new melodies related to previous ones by using Markov processes [8]

and in 1963, Gill's algorithm applied classical AI techniques (hierarchical search with backtracking) [9]. Machines were becoming less expensive and more powerful and the field of algorithmic composition grew along, intersecting the field of Artificial Intelligence.

Composing music imitating an existing corpus has been approached quite successfully. For example, Cope's EMI [10] was able to analyze a musical style, extract patterns of short musical sequences and determine how to use them when composing in that style. More recently, with the rise of neural networks, while not being as widespread as other forms of arts that use Generative Adversarial Networks [11] and other methods, we find many examples of systems following this imitating approach. For example, using long short-term memory (LSTM) recurrent network to learn complete musical structures and produce music in that style [12] or networks based on gate recurrent units (GRU) to produce monophonic melodies by learning from a corpus of different styles [13].

On the other hand, composing with no creative input has been a harder problem for computational intelligence, partly because of the difficulty of assessing the results, although some frameworks have been proposed [2,14–16]. Genetic algorithms have proven to be very suitable to produce novel musical content. In 1994, GenJam [17] was an interactive method to create jazz melodies. More examples of interactive genetic algorithms are Vox Populi [18], based on evolving chords, and Darwin Tunes [19] that studied the evolution of short audio loops. There has been a wide variety of automatic musical fitness assessments. Nam and Kim [19] recently proposed a method to create jazz melodies, encoding explicitly the musical variables in the chromosomes and using a fitness function based on the theory of harmony and note patterns. However, evolutionary and developmental (evo-devo) methods, using indirect encoding, although more complex to design, are meant to provide more unpredictable and diverse results. Musical Organisms [20] defined a gene regulatory network and focused on structural modularity, hierarchy and repetition as a common principle of living beings and music to produce musical scores. Formal grammars have been suitable to express this inner structured nature of music, for example to study melody [21], musical structure [22] or harmony in tonal music, even with a generative approach [23]. L-systems [24] in particular have been frequently used as representational models in evo-devo processes due to their simplicity, sometimes with an interactive fitness and studying ways of visualizing music [25,26].

In Section 2 we present an approach of encoding music by means of L-systems and context-free grammars and then compose music by evolution of these formal models. The two representational systems are actually part of different programs developed from the same principles, but adapted for the composition of both atonal and tonal music. These mathematical systems, which are able to represent self-similarity encoding compositional and performance properties, can evolve as to produce valid and esthetic music scores and then synthesize them in multiple audio formats.

1.2. Application in the Real World

This work is also concerned with how “human” the production can be considered. While measuring the performance of an assembly robot in a production line is straightforward, assessing the quality of other AI systems, particularly those involving simulated cognition or creativity, is always controversial. To date, the Turing Test (TT) [27] is still considered the best way to determine if an artificial system involves some sort of thought. The validity of this approach stands, even if the TT was epistemic rather than ontic or more heuristic than demonstrative [28]. That is the reason why a wide range of variations have been proposed to measure the success of AI systems in related areas. Some of them consider the term “creative” in a very restrictive sense, like the Lovelace Test (LT) [29] that demands full creativeness on the machine under study, which is hard to meet (and verify) for an artificial system. Some interesting and more practical variants are the Musical Directive Toy Test (MDtT) and the Musical Output Toy Test (MOtT) [30], with the second one focusing more on the results of the composition and paying less attention to how the composers were commissioned. Both the MDtT and MOtT remove the essential component

of natural language discourse, while keeping the blind indistinguishability of the TT. As such, they must be considered surveys of musical judgments, not a measure of thought or intelligence.

Another important aspect of generative music systems is their lack of experimental methodology [31] and the fact that typically there is no evaluation of the output by real experts. Delgado [32] considered a compositional system based on expert systems with information about emotions as input and described an interesting study where the extent to which the listeners match the original emotions is evaluated, but the test was only completed by 20 participants. Monteith [33] presented another system and tested it in an experiment where the participants had to identify the emotions evoked for each sample, as well as assess how real and unique they sounded as musical pieces. There were two major limitations in this case: the sample contained only 13 individuals and they knew in advance that there were computer-generated pieces. A more recent example was described by Roig [34], with a questionnaire including three sets of compositions, each with two compositions made by human and the other generated by their machine learning system, which the participants had to identify. In that experiment, there was a bigger sample (88 subjects), however it suffered from lack of control, since the test was performed through the web, and simplicity, because the participants only had to answer which of the three samples was computer generated.

Affective algorithmic composition [35] has promoted new strategies of evaluating the results of automatic composition systems, as compared to human produced music. Here, emotions are understood as relatively short and intense responses to potentially important events that change quickly in the external and internal ambience (generally of social nature). This implies a series of effects (cognitive changes, subjective feelings, expressive behavior or action tendencies) occurring in a more or less synchronized way [36]. This theory of musical emotion [37] asserts that music can elicit emotions from two different involuntary mechanisms of perception and interpretation of the human being: (a) the (conscious or unconscious) assessment of musical elements; and (b) the activation of mental representations, related to memories with significant emotional component [38,39]. Even though self-report has been associated with some obstacles, like a quantitative demand of lexicon or the presence of biases, it is considered one of the most effective ways of identifying and describing the elicited emotions [36]. This work has opted for an open-answer modality of the cited method, since the fact that establishing explicit emotional dimensions might itself condition their appearance due to the human empathy [36].

In Section 3.2 we describe a MDtT designed to assess the compositional and synthesis capabilities of the developed system. The experiment involved more than 200 participants, professional musicians and non-musicians, who listened to and assessed two different pieces of music: one from our generator of tonal music and another composed by a musician. As compared to previous works, the main contributions here are the rigor of the study and the significance of the results.

The conclusion of this endeavor is Melomics, a tool that is able to produce objectively novel and valuable musical scores without using any knowledge of pre-existing music, but trying to model a creative process.

2. Melodies and Genomics

In order to produce full music compositions through evolutionary processes, we need to encode the structure of the music, not only the single pitches and other music variables, but also the relations among them. That meaning, there must be a higher-level structure more compact than the entire composition. In a sense, it is a compressed version of the composition, which implies the presence of some kind of repetitions and structured behaviour, since truly random data would not be susceptible to compression.

While computers are capable of sound synthesis and procedural generation of sound, we are interested in the production of a music score in the traditional staff notation. Some systems aim to reproduce a particular style of a specific artist, period, or genre by using a

corpus of compositions from which recurring structures can be extracted. We pursue music composition from scratch, modeling a creative process able to generate music without imitation of any given corpus. The ability to create music in a specific style is enforced by a combination of two important constraints:

- The encoding, which generates a bias in the search space. That is, by changing the way we represent music we also change what is easy to write and what is difficult -or even impossible- to express. By forcing music to be expressed as a deterministic L-system or a deterministic context-free grammar, regular structures are easier to represent and will appear more often than completely unrelated fragments of music. A set of rules can appear not to provide any repeated structure, but its product might contain a high amount of repetitions and self-similar structures. This is, indeed, an essential aspect in both biology and music [40]. The structure of an entire piece of music is unequivocally encoded, including performing directions, and deploys from a series of derivations. This provides a way to generate repetitions in the product from a simple axiom, without being repetitive.
- The fitness function, which associates a measure of quality to each composition, determined by looking at some of its high-level features, such as duration, amount of dissonance, etc.

The dual effect of having an encoding that restricts the search into a more structured space, where individual changes in the genome can produce a set of organized alterations on the phenotype, combined with the filtering of a fitness function, helps the generation process to “converge” to music respecting a particular style or set of conditions, without any imitation.

2.1. Representational Models

2.1.1. Atonal Music

The atonal system uses an encoding based on a deterministic L-system (genetic part) and some global parameters used during the development phase (epigenetic part). An L-system is defined as a triple (V, S, P) where V is the alphabet (a non-empty, finite set of symbols), $S \in V$ is the axiom or starting symbol and $P \subseteq V \times V^*$ is the set of production rules in the form $A \rightarrow x, A \in V, x \in V^*$. Each production rule rewrites every appearance of the symbol A in the current string into x . Because it is deterministic, there is only one rule $A \rightarrow x$ for each symbol $A \in V$. In our model, we divide symbols of the alphabet V into two types:

- Operators, which are reserved symbols represented by a sequence of characters with the form $@i, i \in \mathbb{Z}$ or $\$j, j \in \{1 \dots m\}$. There are no rules explicitly written for them, the rule $A \rightarrow A$ is assumed instead. They control the current state of the abstract machine, modulating the musical variables: pitch, duration, onset time, effects, volume, tempo, etc.
- Instruments, with the form $\#k, k \in \{0 \dots p\}$. These symbols have an explicit production rule with them on the left side and can represent either a musical structural unit (composition, phrase, idea, etc.) or, if it appears in the final string, a note played by the instrument that has been associated to that symbol.

Since the rewriting process could be potentially infinite, in order to stop it, we introduce the following mechanism: each of the production rules have an associated value r_i , indicating the possibility of that rule to be applied in the next rewriting iteration. There is a global parameter T which serves as the initial value for all r_i and there is a weight I_i for each production rule, to compute the new value of r_i from the previous time step, $r_{i_t} = f(r_{i_{t-1}}, I_i)$. All r_i will progressively get a lower value until reaching 0, when the associated rule will not be applied anymore. The formula is applied if that rule is used at least once in the current iteration. To illustrate the process, let us introduce some of the operators that we use:

\$1 increases pitch value one step in the scale.
 \$2 decreases pitch value one step in the scale.
 \$5 saves current pitch and duration values in the *pitch stack*.
 \$6 returns to the last value of pitch and duration stored in the *pitch stack*.
 \$7 saves current time position in the *time stack*.
 \$8 returns to the last time position saved in the *time stack*.
 \$96 applies dynamic mezzo-forte.
 @60\$102 applies the tempo: quarter equals 60.

Now, let us consider the following simple and handmade L-System

$Ga = (V, S, P)$
 $V = \{\#0, \#1, \#2, \#3, \#4, \$1, \$2, \$5, \$6, \$7, \$8, \$96, @60\$102\}$
 $S = \{\#0\}$

And P consisting of the rules:

$P = \{$
 $\#0 \rightarrow @60\$102\$96\#2\$1\$1\$1\$1\$1\$1\#2$
 $\#1 \rightarrow \#1$
 $\#2 \rightarrow \$7\#3\$8\#4$
 $\#3 \rightarrow \#3\$2\#3\$1\$1\#1\#3$
 $\#4 \rightarrow \#4\#1\$5\$1\$1\$1\#4\$6\#4$
 $\}$

Below we show the whole developmental process to get the composition generated by Ga , considering $T = 1$ and $I_i = 1$ for all the rules. And the following global parameters:

Scale: *C major*
 Tempo: 80 bpm
 Default duration: *quarter note*
 Default dynamic: *mezzo-piano*
 Initial pitch: *middle C*
 Instruments: piano (0), rest (1), church organ (2 and 3), cello (4)

The vector r will show the current values of r_i for all the production rules.

Iteration 0

With no iterations, the resulting string is the axiom, which is interpreted as a single note, played by its associated instrument, the piano, with all the musical parameters being on their default values still.

String: #0

$r = [11111]$

To better illustrate the interpretation procedure, Figure 1 shows the resulting score from the current string, if we supposed that the rewriting counter is 0 for the first rule ($r_0 = 0$) at this point. Since it is not the case, the rewriting process will continue.

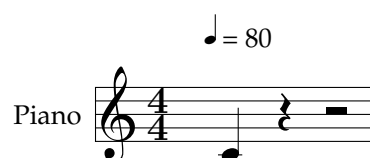


Figure 1. Resulting score from Iteration 0 (see Audio S1).

Iteration 1

For the first iteration, the rule associated to the only symbol in the string needs to be applied and then, applying the formula, r_0 will be set to 0.

String: @60\$102\$96#2\$1\$1\$1\$1\$1\$1\$1#2

$r = [01111]$

This is interpreted from left to right as: change the tempo to “quarter equals 60”, apply the dynamic mezzo-forte, play the current note on instrument 2, rise the pitch seven steps on the given scale and finally play the resulting note on instrument 2. As before, Figure 2 shows the resulting score.



Figure 2. Resulting score from Iteration 1 (see Audio S2).

Iteration 2

The previous score showed the hypothetical outcome if we supposed $r_2 = 0$ already. Since that symbol is in the string and $r_2 = 1$, the rule has to be applied once.

String: @60\$102\$96\$7#3\$8#4\$1\$1\$1\$1\$1\$1\$7#3\$8#4

$r = [01011]$

Figure 3 shows the hypothetical score for Iteration 2. At this point, symbol #2, as opposed to what would have happened if it had stopped at Iteration 1, no longer acts as a playing instrument, but rather as a compositional block that gives place to two instruments. These instruments play at the same time due to the use of the *time stack* operators \$7 and \$8. The “synchronization” between the two instruments is an emerging property from the indirect encoding.

Figure 3. Resulting score from Iteration 2 (see Audio S3).

Iteration 3

The rules with #3 and #4 need to be applied, resulting:

String: @60\$102\$96\$7#3\$2#3\$1\$1#1#3\$8#4#1\$5\$1\$1\$1#4\$6#4\$1\$1\$1\$1\$1\$1\$7

#3\$2#3\$1\$1#1#3\$8#4#1\$5\$1\$1\$1#4\$6#4

$r = [01000]$

The music has acquired a structure, the initial rewriting resulted in a change of tempo and dynamic and a shift up in the pitch one octave for the second part of the composition. The following rewriting developed two instruments playing in polyphony and the latest one provided the final melody to each instrument.

There is a fourth iteration that will rewrite symbol #1 into itself and set $r_1 = 0$. The score shown in Figure 4 for Iteration 3 is the same as the final score in this case.



Figure 4. Resulting score from Iteration 3 (see Audio S4).

Supplementary material Scores S10 shows two of the first actual pieces that were generated with the atonal system.

2.1.2. Tonal Music

While the system used to produce atonal music can also generate tonal music, in order to favor the emergence of structures and elements that are usually present by convention in popular Western music, there are different changes introduced to the original system.

In order to establish a compositional structure as it is common in a traditional composing process, inspired by Bent's Analysis [41], the production rules are explicitly structured in five hierarchical levels:

- **Composition.** This is the most abstract level and it is formed by a sequence of similar or different kinds of periods, possibly with music operators (alterations in tone, harmony, tempo, macro-dynamics. . .) between each of them.
- **Period.** This is the highest structural subdivision of a composition. There can be more than one type of period, built independently, becoming separate musical units recognizable in the composition.
- **Phrase.** This is the third structural level, the constituent material of the periods.
- **Idea.** Constitutes the lowest abstract level in the structure of a composition. A phrase can be composed by different ideas that can be repeated in time, with music operators in the middle. A musical idea is a short sequence of notes generated independently for each role, using many different criteria (harmony, rhythm, pitch intervals, relationship with other roles. . .).
- **Notes.** This level is the most concrete level, composed only by operators and notes played by instruments.

Global parameters are introduced to force the occurrence of tonal music. Some of these are applied straight away when creating a new genome, by filtering the use of certain strings as the right hand side of a production rule and some others are used as part of the fitness function (see Section 2.2.2 below). These parameters can be grouped in the following categories:

- **Duration.** Boundaries or specific values to set up the duration or length of the composition or any of the structural levels.
- **Tempo.** Specific values or boundaries to establish tempos along a composition.
- **Dynamics.** Specific values or boundaries to establish the macro-dynamics to be applied along a composition.
- **Role types.** List of roles or behaviors that may appear in the compositions. There are currently 89 different roles implemented that can be classified in melody, accompaniments, homophony (harmonized melody) and counterpoint. The harmonic and rhythmic accompaniments can appear in different forms: Chords, Bass, Pads, Arpeggios, Ostinati, Drums and Percussion.
- **Arpeggio configuration.** If the role arpeggio is enabled, some additional parameters must be given, such as mode or contour type, durations or time segments to divide the

primary rhythmic cell, scope or number of octaves and tie notes, indicating whether consecutive similar notes shall be tied or not.

- Instruments. For each role there is a list of allowed instruments, plus additional specific properties and their probability to be chosen.
- Rhythmic incompatibilities. Defines the incompatibilities of instruments to play notes at the exact same moment.
- Scales. A set of modes and notes of reference to build the musical scales to be used in the compositions. Resulting scales can be C-major, C-minor, D-Dorian or C-Mixolydian.
- Harmony. This category includes parameters to describe how to build the harmony in the different compositional levels; containing for example the allowed chords and roots, a measure of the allowed dissonance between melodies or the form of the chord progressions.
- Rhythmic modes. Includes the allowed types of measures, the types of accents to perform the notes and other related parameters.
- Rhythmic patterns. For each role a list of valid note values or patterns may be given, establishing the desired frequency of occurrence in a composition.
- Melodic pitch intervals. For each role of the melody kind, a weighted list of pitch intervals may be given to define how the melodic contour should be.
- Texture. This sub-section allows the possibility to define rules for the inclusion of roles; different types of dependencies between them; the compositional units where they are forced, allowed or prohibited to appear in; and general evolution of the presence of instruments.

There are a few new operators to improve harmony management, some of them with a more complex interpretation process and dependencies with other operators and global parameters, like the operator M to create chords (see development example below).

Since we enforce a strict hierarchical structure, it is useful to distinguish the symbols that represent compositional elements that will be rewritten during the development from those that have the same specific meaning at any given moment. For this reason we use an encoding based on deterministic context-free grammar, defined as (V, Σ, S, P) , where V is the set of non-terminal symbols, Σ is the set of terminal symbols (the alphabet), $S \in V$ is the axiom or starting symbol, $P \subseteq V \times \{V \cup \Sigma\}^*$ is the set of production rules. The symbols in the set V can be identified with the structural units: composition, periods, phrases and ideas, while the symbols in Σ represent notes and music operators, such as modulators of pitch, duration, current harmonic root or current chord. In our implementation, on the right hand side of the production rules, there are only terminal symbols or non-terminal symbols from the following level in the hierarchy in a decreasing order.

To illustrate the rewriting process, let us introduce some of the reserved terminal symbols and their interpretation:

N increases the counters pitch and harmonic root in one unit.

n decreases the counters pitch and harmonic root in one unit.

$[$ saves in a stack the current value of pitch, harmonic root and duration.

$]$ restores from the stack the last value of pitch, harmonic root and duration.

$<$ saves in a stack the current time position, value of pitch, harmonic root and duration.

$>$ restores from the stack the last saved time position, value of pitch, harmonic root and duration.

$W4.0$ applies the macro-dynamic mezzo-forte.

$M0.0.0.0$ makes the next symbol linked to an instrument to play the root note of the current chord, instead of the current pitch.

Let us define Gt , a simple and handmade grammar to illustrate the development using this new model:

$$\begin{aligned}
 Gt &= (V, \Sigma, S, P) \\
 V &= \{Z, A, B, C, D, E, F\} \\
 \Sigma &= \{N, n, [,], <, >, W4.0, a, b, s, M0.0.0.0\} \\
 S &= \{Z\}
 \end{aligned}$$

And P consisting of the following rules:

$$\begin{aligned}
 P &= \{ \\
 Z &\rightarrow W4.0[ANNNNNNNA]B \\
 A &\rightarrow CC \\
 B &\rightarrow D \\
 C &\rightarrow ENEnE \\
 D &\rightarrow FFNF \\
 E &\rightarrow < anaNNsa > M0.0.0.0bM0.0.0.0bsM0.0.0.0b \\
 F &\rightarrow as[NNNa]a \\
 &\}
 \end{aligned}$$

Iteration 0 (composition)

String: Z

The axiom Z represents the highest structural level, the composition.

Iteration 1 (periods)

String: $W4.0[ANNNNNNNA]B$

The composition develops into two identical musical units (periods), represented by the symbol A , separated by seven steps in the pitch dimension and then followed by the period B .

Iteration 2 (phrases)

String: $W4.0[CCNNNNNNNCC]D$

The two types of periods A and B would develop into two simple sequences of phrases, CC and D respectively.

Iteration 3 (musical ideas)

String: $W4.0[ENEnEENEnENNNNNNNNENEnEENEnE]FFNF$

Each phrase C and D consists of a sequence of three ideas.

Iteration 4 (notes)

String : $W4.0[$

$< anaNNsa > M0.0.0.0bM0.0.0.0bsM0.0.0.0bN < anaNNsa > M0.0.0.0bM0.0.0.0bsM0.0.0.0bn$
 $< anaNNsa > M0.0.0.0bM0.0.0.0bsM0.0.0.0b$
 $< anaNNsa > M0.0.0.0bM0.0.0.0bsM0.0.0.0bN < anaNNsa > M0.0.0.0bM0.0.0.0bsM0.0.0.0bn$
 $< anaNNsa > M0.0.0.0bM0.0.0.0bsM0.0.0.0b$
 $NNNNNNN$
 $< anaNNsa > M0.0.0.0bM0.0.0.0bsM0.0.0.0bN < anaNNsa > M0.0.0.0bM0.0.0.0bsM0.0.0.0bn$
 $< anaNNsa > M0.0.0.0bM0.0.0.0bsM0.0.0.0b$
 $< anaNNsa > M0.0.0.0bM0.0.0.0bsM0.0.0.0bN < anaNNsa > M0.0.0.0bM0.0.0.0bsM0.0.0.0bn$
 $< anaNNsa > M0.0.0.0bM0.0.0.0bsM0.0.0.0b]$
 $as[NNNa]aas[NNNa]aNas[NNNa]a$

The last idea, F , is interpreted only by the instrument linked to a while the idea E is performed by the instruments associated to a and b , in polyphony; the latter always playing a harmony consisting of the root note of the current chord. To interpret the final string we will use the following values:

Initial scale: *C major*
 Tempo: 80 bpm
 Default duration: *quarter note*
 Default dynamic: *mezzo-piano*
 Initial pitch: *middle C*
 Initial chord: *major triad*
 Initial root: *I*
 Instruments: violin (symbol *a*), double bass (symbol *b*), musical rest (symbol *s*)

See the resulting composition in Figure 5.

Supplementary material Theme S11 provides the genome, the auxiliary MIDI and the MP3 of an actual piece for clarinet generated with the tonal system.

♩ = 80

The figure displays four systems of musical notation. Each system consists of two staves: a Violin (Vln.) staff in treble clef and a Double Bass (Db.) staff in bass clef. The time signature is 4/4. Above the first system, a tempo marking indicates a quarter note equals 80 beats per minute (♩ = 80). The notation includes various note values (quarter notes, eighth notes, sixteenth notes) and rests. The first system shows a simple melody in the violin and a bass line in the double bass. The second and third systems show more complex, rhythmic patterns in both parts. The fourth system shows the violin part continuing with a simple melody while the double bass part is mostly rests.

Figure 5. Resulting score from Iteration 4 (see Audio S5).

2.2. Composing Process

Melomics uses a combination of formal grammar to represent music concepts of varying degrees of abstraction and evolutionary techniques to evolve the set of production rules. The system can be used to compose both atonal and tonal music and, although using slightly different encoding methods and a much stronger set of constraints in the second case, they both share a similar structure and execution workflow (see Figure 6). From a bio-inspired perspective this can be thought of as an evolutionary process that operates over a developmental procedure, held by the formal grammar. The music “grows” from the initial seed, the axiom, through the production rules to an internal symbolic representation

(similar to a MIDI file), where finally the compositions are subject to test by the set of constraints provided. The execution workflow can be described as follows:

- Filling the desired input parameters. These parameters represent musical specifications or directions at different levels of abstraction, such as instruments that can appear, amount of dissonance, duration of the composition, etc., with no need for creative input from the user.
- The initial gene pool is created randomly, using part of the input parameters as boundaries and filters.
- Each valid genome, based on deterministic grammar and stored as a plain text file, is read and developed into a resulting string of symbols.
- The string of symbols is rewritten after some adjustment processes that can be of different forms: cleaning of the string, for example by removing sequences of idempotent operators; adjustments due to physical constraints of the instruments, like the maximum number of simultaneous notes playing; or the suppression (or emergence) of particular musical effects.
- Each symbol in the final string has a musical meaning with a low level of abstraction, which is interpreted by the system through a sequential reading from left to right and stored in an internal representation.
- Once again the musical information will be adjusted and stabilized. For example, shift the pitches to satisfy constraints in the instruments' tessituras, discretization of note durations and so on.
- The input directions are used to assess the produced composition that might be discarded or pass the filter. In that case it is saved as a 'valid' composition.
- A discarded theme's genome is replaced by a new random genome in the gene pool. On the other hand, a "valid" genome is taken back to the gene pool, after being iteratively subject to random mutations and possible crossover with another genome in the gene pool (see Section 2.2.1), until it passes the filters defined at genome level (the same as with a random genome).
- If desired, any composition (usually the ones that pass the filter) can be translated, using the different implemented modules, to standard musical formats that can be symbolic (MIDI, MusicXML, PDF) or audio (WAV, MP3, etc.), after executing a synthesis procedure with virtual instruments, which is also led by the information encoded in the composition's genome.

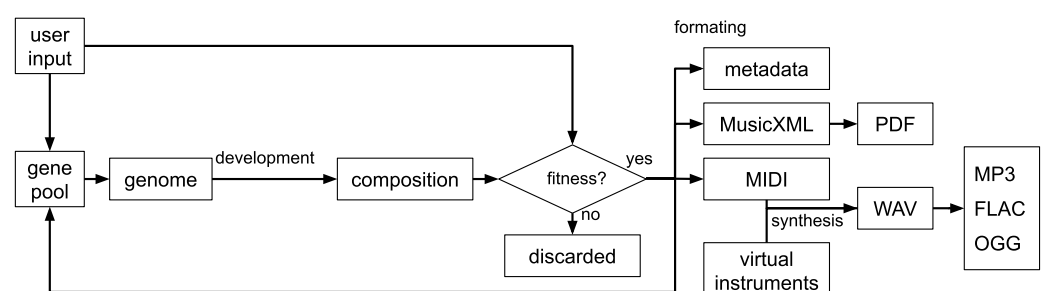


Figure 6. A graphical representation of the composition workflow.

The average time to build a new valid genome varies very much depending on the style and constraints imposed at this level. Using single core on an Intel Xeon E645, for a very simple style it can be ready in less than one second, while for the most complex style, symphonic orchestra in the atonal system, the process takes 283 s on average. Given a built gene pool, the time to obtain a new valid composition, which includes executing the developmental process possibly repeatedly until passing the filters, varies a lot too. It can take 6 s for the simplest atonal style, while it takes 368 s on average for the most complex style in the tonal system.

2.2.1. Mutation and Crossover

The genomes can be altered in any way possible in the valid set of symbols defined by removing, adding or altering symbols in any position; the developmental procedure will always give place to a valid musical composition. However, the disruption provided by genetic operators (mutation and crossover) should be balanced in order for the system to converge properly to the given directions. In Melomics, a small amount of disruption combined with the hierarchical structure granted by the encoding makes the system produce reasonable results since early iterations, typically less than 10.

For atonal music we allow stronger alterations, since the styles addressed in general are less constraining than the popular styles pursued with the tonal system. The implemented mutation are: (1) changing a global parameter (e.g., default tempo or dynamics); (2) changing the instrument associated to a symbol in the grammar; (3) adding, removing or changing a symbol on the right hand side of a production rule; (4) removing a production rule; and (5) adding a new production rule, copying an existing one or generated randomly, and introducing the new symbol randomly in some of the existing production rules. The implemented crossover mechanism consists of building a new genome taking elements from two others. The values for the global parameters and the list of instruments are taken randomly from any of the source genomes. The set of production rules is taken from one of the parent genomes and then the right hand side can be replaced with material from the other parent. The disruption introduced with these mechanisms are still too high, even for atonal music, hence the mutation operations (4) and (5) are used with less probability and the rule with the symbol associated to the musical rest is always kept unaltered. Audio S6 shows a mutation of the Nokia tune (https://www.youtube.com/results?search_query=nokia+tune, accessed on 30 April 2021), reverse engineered into the system, where the instrument has been altered as well as the rules at the lowest level of abstraction, resulting in some notes being changed while maintaining the more abstract structure.

For tonal music the mutation and crossover operations are similar, but more restricted and executing less of them at each iteration. Mutations allowed are: (1) changing a global parameter; (2) changing an instrument for another valid for the same role; (3) adding, removing or changing a non-terminal symbol or a terminal reserved symbol (operators, no instruments associated) on the right hand side of a production rule, provided that it does not alter the five-level hierarchical development; and (4) removing or adding a new terminal non-reserved symbol (Instrument), in the second case duplicating appearances and forcing polyphony with an existing one (enclosing the new one with the symbols <, > and placing it to the left) and assigning an instrument of the same role. The crossover is also similar, taking the genome of one of the parents and only replacing a few rules with material of the same structural level from the other. The global parameters are taken randomly from any of them and the same for the instruments, respecting role constraints. Audio S7 shows a sample generated in the style *DiscoWow2* and Audio S8 shows a crossover of it with the tune SNSD SBS Logo Song (https://www.youtube.com/results?search_query=SNSD+SBS+Logo+Song, accessed on 30 April 2021) of Korean TV, reverse engineered into the tonal system.

2.2.2. Fitness

Both the atonal and the tonal systems count on a set of parameterized rules to guide the composing process, by allowing the development of those composition that comply with the rules, while filtering the rest of them out. There are (a) global rules that basically constitute physical constraints of the musical instruments, such as the impossibility for a single instrument to play more than a certain number of notes simultaneously or to play a note too short or too long; and (b) style based constraints that encode expert knowledge and are used to assure the emergence of a particular kind of music, resembling the way a human musician is requested to create music in a certain style. This latter kind of rules, in general looser in the atonal system, can be grouped as follows:

- **Duration.** For example lower and upper boundaries for the duration of a composition, which exist both in the atonal and tonal systems and are assessed at the end of the developmental process, on the phenotype, when the musical information is explicitly written in the internal symbolic format. There are also duration filters for other compositional levels depending on the style.
- **Structure.** These are constraints to the number and distribution of compositional blocks.
- **Texture and dynamics.** In the atonal system both polyphonic density and global dynamics are checked at every point of the musical piece, according to the specified function at the input, with a margin of tolerance. For the tonal system the procedure is more restrictive and applied at genotype level. Valid musical roles, dynamics and instrumental density are defined for each type of compositional block.
- **Instruments.** Both systems check the instruments chosen to build a valid genome in a defined style.
- **Harmony and rhythm.** In the atonal system there is a measure of the amount of dissonance, a count of pitch steps and a count of changes of note duration. Each of these properties has a lower and an upper threshold that must be complied for a specified time window that moves along the composition. The tonal system is more restrictive. For each defined style there is the list of optional parameters described in Section 2.1.2, where some of them translate directly to filters of the fitness function. The main filters regarding harmony and rhythm are: a set of valid modes and tones; valid measure types; valid rhythmic modes for the melody; valid chords for each role type; valid “harmonic transitions”, as a way to assess and filter out certain chord progressions; and valid rhythmic patterns, expressed in terms of the current measure type and including rests. All the filters have a tolerance value associated and most of them can be defined globally or at the level of compositional block (period, phrase or idea).

3. Results

3.1. Preliminary Assessments

We used music information retrieval tools to measure similarity of compositions generated under common musical requirements, then compare these themes with other generated in a different style and also with music created by humans.

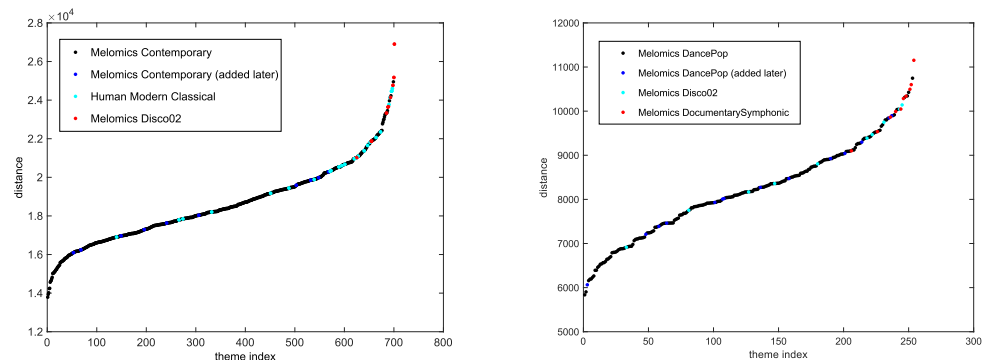
To deal with music in a symbolic way, we used the open-source application jMIR [42], in particular jSymbolic, which is able to extract information from MIDI files. We set up the tool to get features classified into the categories of instrumentation, texture, rhythm, dynamics, pitch statistics, melody and chords adding up to 111 in total.

For the atonal system, we picked a collection of 656 pieces of classical contemporary music, with different ensembles, then computed the average value for each feature, obtaining a sort of centroid of the style. After that, we added to the bundle 10 more pieces from the same contemporary classical substyle, 10 from the style *Disco02*, produced by the tonal system and 25 pieces from the Bodhidharma dataset [43] tagged as “Modern Classical”, created by human composers, representing 701 pieces in total.

For the tonal system we performed a similar test. Choosing 220 themes from Melomics’s *DancePop* and excluding 14 themes to compute the centroid. Then we added: these 14 pieces, 10 from the style Melomics’s *Disco02* that we consider the predecessor of *DancePop* and 10 more from Melomics’s atonal *DocumentarySymphonic*.

For both cases, Figure 7 shows a representation of the distances of each theme to the computed centroid. In the atonal, the extra pieces generated of contemporary music appear scattered around the centroid of the group that had been generated previously (mean and standard deviation: $\mu = 1819.7$, $\sigma = 1603.2$ and $\mu = 1842.54$, $\sigma = 1838.3$ respectively); the human made compositions in similar styles appear close but shifted to farther distances ($\mu = 2101.88$, $\sigma = 2158.76$), while music created with different directions appear at the farthest distance to the centroid ($\mu = 2363.76$, $\sigma = 1755.02$). In the second case, the pieces of the same style that were added later are close to the centroid of the group of reference

($\mu = 8091.64$, $\sigma = 977.71$ and $\mu = 8273.57$, $\sigma = 1071.6$); compositions of a similar style appear close but shifted farther ($\mu = 8879.0$, $\sigma = 1065.41$) and music of a more different style are located at the farthest distance ($\mu = 10,141.4$, $\sigma = 575.53$).



(a) Study of distances for *Contemporary* style

(b) Study of distances for *DancePop* style

Figure 7. Chart (a) shows a representation of the distances to the centroid of Melomics's *Contemporary* style from themes of other styles configured with Melomics and a collection of themes composed by human artists from the Bodhidharma dataset. Chart (b) shows the distances to the centroid of Melomics's *DancePop* style from themes of different styles configured with Melomics tonal system.

We also used the API provided by The Echo Nest [44] (acquired by Spotify in 2014), to put in context Melomics music, this time analyzing audio files, among another set of musical pieces, natural sounds and noise, extracting some of the musical properties that they define (loudness, hotness, danceability, energy) and obtaining similar results.

3.2. Experiment in the Real World

The aim, through an MDtT design, is quantifying to what extent are Melomics and human music perceived as the same entity and how similarly do they affect the emotions of a human listener (musician or naive). The hypothesis is that a computer-made piece is indeed equivalent to conventional music in terms of the reactions elicited in a suitable audience.

In an effort to verify this hypothesis, an experiment registered the mental representations and emotional states elicited in an audience while listening to samples of computer-generated music, human-generated music and environmental sounds. It involved 251 participants (half of them professional musicians) who reported on the mental representations and emotions that were evoked while listening to the various samples. The subjects were also asked to differentiate the piece composed by computer from the one created by a human composer.

3.2.1. Methodology

Two musical pieces were commissioned in a specific style to both a human composer and the Melomics tonal system. In the first stage a comparison between both pieces was made using an approach that focused on human perception: listeners were asked what mental images and emotions were evoked on hearing each sample. In this phase of the experiment the listeners were not aware that some of the samples had been composed by a computer. In an effort to gauge whether or not both works were perceived in a similar way, the nature of the composer was not revealed until the second stage.

Data analysis has been performed with a fourfold contingency table (two by two) and the differences between groups were evaluated by the chi-squared test with continuity correction. The Fisher test has been used only in those cases where the expected frequency was lower than 5. The significance level has been established at $p < 0.05$. In the second phase of the study, the sensitivity (i.e., the capacity to correctly identify the pieces composed by the computer) and the specificity (i.e., correct classification of pieces composed by humans) were also evaluated for both musicians and non-musicians, at a confidence level of 95%.

3.2.2. The Audio Samples

The specifications for the musical pieces were kept simple and the duration short, in order to ease their assessment, especially by non-musician participants:

style: guitar ballad

instruments: piano, electric piano, bass guitar and electric guitar

bar: 4/4

BPM: 90

duration: 120s

structure: A B Ar

scale: major

Guitar ballad was chosen also because its compositional rules were already coded within the system. The final audio files were obtained in MP3 format with a constant bitrate of 128 kbps and shortened to a duration of 1 m 42 s, meaning the beginnings and endings of the pieces (around 9 s each) were removed since they typically render constant composing patterns in popular music. Following composition, both the human- and the computer-composed ballads were doubly instantiated by means of, on the one hand, performance by a professional player and, on the other hand, through computerised reproduction. For the purposes of the performance the human player interpreted the scores of both works while the computer automatically synthesized and mixed a MIDI representation of both pieces using a similar configuration of virtual instruments and effects, the four combinations resulting in corresponding music samples. Table 1 shows how the pieces were labelled: HH stands for human-composed and human performance, CC for computer-composed and synthesized, HC for human-composed and computer-synthesized, CH for computer-composed and human performance, and NS for natural sounds. In contrast to the musical samples, this final sample has been introduced in an effort to gauge the listeners' response to non-musical sounds and consists of a two-minute excerpt from natural sounds (Jungle River, Jungle birdsong and Showers from The Sounds of Nature Collection [45] combined with animal sounds [46]. Audio S9 contains the five audio samples used in the study.

Table 1. Classification of audio samples according to the composer and the interpreter.

		Composed		
		Human	Computer	Nature
interpreted	human	HH	CH	-
	computer	HC	CC	-
	nature	-	-	NS

3.2.3. Participants

The experiment was carried out in two facilities in Malaga (Spain): the Museum for Interactive Music and the Music Conservatory. Subjects were recruited via posters in the museum facilities and internal calls among students and educators at the conservatory. Selected participants ranged in age from 20 to 60 and the answers were processed differently according to music expertise: subjects with five or more years training were labelled as 'musicians' while subjects with less or no training were classified as 'non-musicians'. Musicians are assumed to process music in a far more elaborate manner and possess a wider knowledge of musical structure, so they would be expected to outperform non-musicians in a musical classification task.

The final sample consisted of 251 subjects, the mean age being 30.24 ($SD = 10.7$) years, with more musicians ($n = 149$) than non-musicians ($n = 102$). By gender the sample featured marginally more women ($n = 127$) than men ($n = 124$) and, in terms of nationality, the majority of the participants were Spanish ($n = 204$), the remainder being from other, mainly European countries ($n = 47$).

3.2.4. Materials

In both facilities the equipment used to perform the test was assembled in a quiet, isolated room in order to prevent the participants from being disturbed. Each of the three stations used for the experiment consisted of a chair, a table and a tablet (iPad 2). The tablet contained both the web application that provided the audio samples and the questionnaire that was to be completed by means of the touch screen. The text content (presentation, informed consent, questions and possible answers) were presented primarily in Spanish with English translations located below each paragraph. The device was configured with a WIFI connection to store the data in a cloud document and featured headphones (Sennheiser EH-150) both for listening to the recordings and to avoid external interferences. The iPad device was configured to perform only this task and users were redirected to the home screen when all the answers had been saved. The possibility of exiting the app was also disabled. The test is publicly available at <http://www.geb.uma.es/mimma/>, accessed on 30 April 2021.

3.2.5. Procedure

Supported by neuroscientific studies [47], critiques to computer-made compositions are suitable to be affected by anti-computer prejudice, if knowing in advance the non-human nature of the author. Hence, during the test, each subject was informed that they were undergoing an experiment in music psychology, but the fact that it involved computer-composed music was not mentioned at the beginning so as not to bias the results [48]. Participants were also randomly assigned to one of two different groups and the compositions were distributed between these two groups in such a way that each subject listened to the musical pieces as rendered by the same interpreter (Table 2). In this way, the responses were independent of the quality of the execution since each subject would listen to both the human and computer compositions interpreted either by the artist or by the computer, meaning that potential differences in composition and performance became irrelevant.

Table 2. Distribution of musical pieces into groups according to the interpreter.

	Group A	Group B
Phase I	HH/CH/NS	CC/HC/NS
Phase II	HH/CH	CC/HC

The workflow of the test detailed in Figure 8 shows that the subject is first introduced to the presentation and the informed consent screens and then prompted for personal data relevant to the experiment (see specific questions in Table 3). The subject is then assigned to either group A or B, listens to five sequential audio recordings and answers a number of questions. The listening and answering process is divided into two phases:

1. During Phase I each subject in group A listens to the three pieces in random order. HH and CH have been composed by a human and by our computer system, respectively, and both are performed by a human musician. Subjects assigned to group B proceed similarly, listening to the same compositions, but in this case synthesized by computer. Both groups also listened to the same natural sounds recording (NS). Having listened to each piece the subject is then asked whether the excerpt could be considered music and what mental representation or emotional states it has evoked in them. This final question requires an open answer as the subject was not given a list of specific responses.
2. In Phase II, the subject listens to the same musical pieces (but not the natural sounds), following which they are asked whether they think the piece was composed by a human or by a computer. As previously stated, it is important that identification of the composer is withheld until the second phase so that subjects can provide their

assessment of the music in Phase I without the potential bias of Phase II, in which the subject becomes aware that the music might have been composed by a computer.

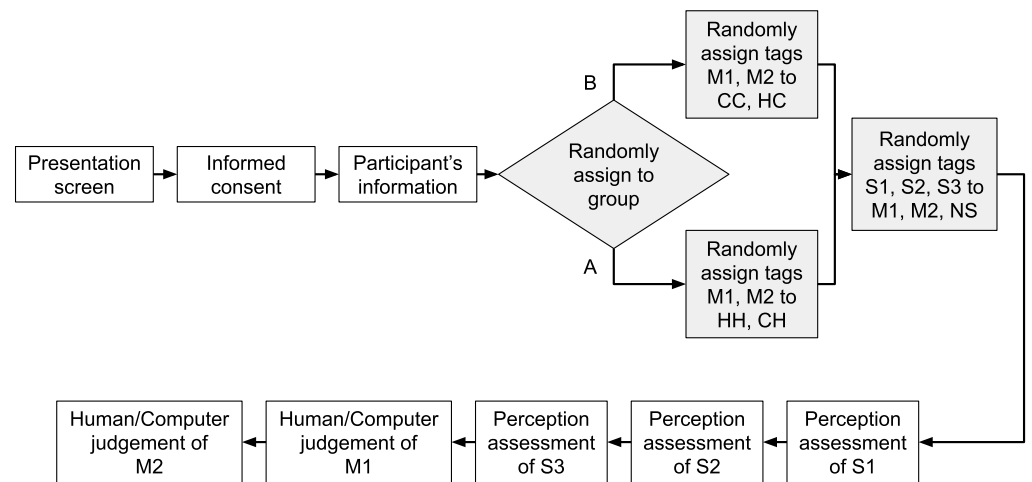


Figure 8. Test workflow. After the initial pages, the participants are randomly assigned to one group (A or B). Then, the samples are played and evaluated in a random order. Finally, the human or artificial nature of the musical pieces is assessed.

Table 3. Information requested to participants and questions to interrogate the personal opinion and feelings elicited by the recordings.

Stage	Questions	Input Type
Participant's information	Age	Numeric
	Sex	Option button
	Country	Text box
	Email address (optional)	Text box
	Do you have at least 5 years of academic musical training?	Option button
Perception assessments	Q1. Would you say that what you have listened to is music?	Option button
	Q2. Does it generate mental representations when you listen to it?	Option button
	Q3. If so, could you write three mental representations that have come to mind listening to it?	Text box
	Q4. Does it generate any feelings?	Option button
	Q5. If so, could you write three feelings that you have felt while listening to it?	Text box
Human/Computer judgements	Q6. Do you think that the piece you have listened to has been composed by a composer or by a computer?	Option button

In summarizing the process, a subject assigned to group A might proceed as follows:

1. Presentation screen
2. Informed consent
3. Participant's information
Perception assessment of S1
4. Listen to HH (labelled as M1, and then S1)

5. Q1-Q5
Perception assessment of S2
6. Listen to NS (labelled as S2)
7. Q1-Q5
Perception assessment of S3
8. Listen to CH (labelled as M2, then S3)
9. Q1-Q5
Human/computer judgement of M1
10. Listen to HH
11. Q6
Human/computer judgement of M2
12. Listen to CH
13. Q6

3.2.6. Experiments Results

Table 4 shows the results in percentage of the *yes/no* questions provided in the first (blind) phase of the experiment. All the questions asked to the subject referred to the piece that had just been listened to: in random order, HH, CH and NS for subjects of group A, and CC, HC and NS for group B. None of the contrasts performed, for the total sample or for each type of listener, was found significant.

Table 4. Results of the *yes/no* questions proposed in Phase I of the experiment. The percentages of “yes” are shown. Q1: Would you say that what you have listened to is music? Q2: Does it generate mental representations when you listen to it? Q3: Does it generate any feelings?

	Sample		Musicians		Non-Musicians	
	Group A (<i>n</i> = 134)	Group B (<i>n</i> = 117)	Group A (<i>n</i> = 89)	Group B (<i>n</i> = 60)	Group A (<i>n</i> = 45)	Group B (<i>n</i> = 57)
	HH	HC	HH	HC	HH	HC
Q1	98.5	96.6	98.9	100.0	97.8	93.0
Q2	70.1	62.4	68.5	60.0	73.3	64.9
Q4	82.1	82.1	84.3	81.7	77.8	82.5
	CH	CC	CH	CC	CH	CC
Q1	97.8	99.1	97.8	100.0	97.8	98.2
Q2	63.4	55.6	66.3	51.7	57.8	59.6
Q4	87.3	80.3	89.9	76.7	82.2	84.5
	NS	NS	NS	NS	NS	NS
Q1	36.6	38.5	41.6	41.7	26.7	35.1
Q2	91.8	95.7	94.4	95.0	86.7	96.5
Q4	83.6	84.6	84.3	81.7	82.2	87.7

An affirmative answer was almost unanimously given to the first question (Q1) by the subjects after listening to a sample of a musical piece (as opposed to NS), independently of who composed or interpreted it. By contrast, the NS sample was classified as non-music, although the results were narrower. Regarding the second question (Q2), all the musical pieces elicited mental images, with global percentages ranging from 55.6% to 70.1% and without significant differences observed in terms of who composed or interpreted the pieces ($p > 0.05$). Meanwhile, natural sounds elicited mental images in more than 90% of the cases, more than with any of the musical samples. Regarding the responses elicited by the recordings (Q4), subjects answered affirmatively (up to 80%) to both musical pieces—regardless of who was the composer or the interpreter—and the natural sounds. Only in

the case of the music sample composed by computer and played by human musicians did the global percentage approach 90%. Most of the answers (179) were given in Spanish. Among the responses provided in English (23), three were given by US citizens, three by UK citizens, one by a Canadian and the remaining sixteen by subjects from non-English speaking countries. One subject provided the answers in German and the remainder (48) left the text boxes empty.

Regarding mental representations, three different categories were established: 'nature/naturaleza', 'self/sí mismo' and 'others/otros'. This taxonomy aims to clearly distinguish whether the representational object is associated with the subjective realm or with the world [49]. Regarding emotions, the model proposed by Diaz & Flores [50] contains 28 polarized categories and an associated thesaurus for the task of grouping and was adopted for the purposes of classification. The aim was, first of all, to provide sufficient classes in which to include most of the participants' subjective characterizations while avoiding oversimplification and, secondly, to reduce the number of classes in the analysis where possible. The particular clustering model can be described as a combination of Plutchik's affective wheel [51], with the terms arranged together based on similarity, and the valence-arousal plane [52] that settles polarities for each of the present terms.

In order to analyse the distribution of frequencies in the assembled descriptive data, the different groups were compared using a chi-squared test with continuity correction in two by two tables, establishing the significance level at $p < 0.05$. Among the 753 descriptive entries for mental representations (251 subjects \times 3 samples), 501 were classified into the categories defined above: nature ($n = 300$), others ($n = 112$) and self ($n = 89$); 23 were not classified and subjects did not offer a response (the box was left empty) in 229 cases. With respect to the emotional states, among the 753 entries, 600 were classified, 9 were not classified and 144 were left empty. Table 5 shows the contingency table for the five audio samples and the 501 classified mental representations, those descriptions relating to 'nature' (59.9%) being the most abundant. In the study of distributions there exist significant differences, since the mental representations elicited by the NS sample are mostly described using terms relating to 'nature' (91.7%). In the rest of the samples (the musical ones) the percentages are distributed more evenly among the categories, with a slight predominance of the type 'others'.

Table 5. Contingency table audio sample, type of mental representation.

			Type of Mental Representation			Total
			Nature	Others	Self	
sample	CC	<i>n</i>	19	21	12	52
		%	36.5	40.4	23.1	100.0
	CH	<i>n</i>	30	24	24	78
		%	38.5	30.8	30.8	100.0
	HC	<i>n</i>	15	29	17	61
		%	24.6	47.5	27.9	100.0
	HH	<i>n</i>	26	36	19	81
		%	32.1	44.4	23.5	100.0
	NS	<i>n</i>	210	2	17	229
		%	91.7	0.9	7.4	100.0
total	<i>n</i>	300	112	89	501	
	%	59.9	22.4	17.8	100.0	

In the study of the described emotions, the 600 valid descriptions of emotional stimuli were classified into 20 of the 28 established categories. The most frequent emotional state was "calm" (57.5%), followed by "sadness" (13.3%) and the rest of them appearing with a notably smaller percentage. Since there are several categories and groups of study, in order to compute the hypothesis testing, they have been arranged into three classes, the

two most frequent emotional states, “calm” and “sadness”, and “other emotions”. Table 6 shows the results. There is a significant difference in the group NS ($p < 0.001$), since most of the emotional descriptions are associated with “calm” (89.7%), while the rest of groups show similar percentages.

Table 6. Contingency table audio sample, affective system (grouped).

			Affective System (Grouped)			Total
			Calm	Sadness	Other Emotions	
sample	CC	<i>n</i>	33	16	40	89
		%	37.1	18.0	44.9	100.0
	CH	<i>n</i>	53	27	32	112
		%	47.3	24.1	28.6	100.0
	HC	<i>n</i>	34	11	47	92
		%	37.0	12.0	51.1	100.0
	HH	<i>n</i>	42	24	37	103
		%	40.8	23.3	35.9	100.0
	NS	<i>n</i>	183	2	19	204
		%	89.7	1.0	9.3	100.0
	total	<i>n</i>	345	80	175	600
		%	57.5	13.3	29.2	100.0

In addition, the emotional states were analysed according to the polarity (or valence) in the 14 defined axes. Table 7 shows the results of this arrangement. The majority of emotional descriptions fit into the ‘pleasant’ category or positive valence (75.7%). Nevertheless, there is once again a significant difference with the group NS ($p < 0.001$), since 96.1% of the descriptions come under the category of ‘pleasant’, while in the music samples this figure ranges from 60.7% to 71.7%.

Table 7. Contingency table audio sample, valence.

			Valence		Total
			Unpleasant	Pleasant	
sample	CC	<i>n</i>	28	61	89
		%	31.5	68.5	100.0
	CH	<i>n</i>	44	68	112
		%	39.3	60.7	100.0
	HC	<i>n</i>	26	66	92
		%	28.3	71.7	100.0
	HH	<i>n</i>	40	63	103
		%	38.8	61.2	100.0
	NS	<i>n</i>	8	196	204
		%	3.9	96.1	100.0
	total	<i>n</i>	146	454	600
		%	24.3	75.7	100.0

Table 8 provides a summary of the final stage of the experiment in which the participants responded to the question regarding the identity of the composer of each of the two musical pieces they had previously listened to. These questions were again presented in random order (HH, CH for group A subjects and HC, CC for group B).

Table 8. Results of the explicit inquiry about the nature of the composer in Phase II of the experiment.

Q6	Sample (<i>n</i> = 251)					Musicians (<i>n</i> = 149)					Non-Musicians (<i>n</i> = 102)				
	Group A		Group B		<i>p</i>	Group A		Group B		<i>p</i>	Group A		Group B		<i>p</i>
	<i>n</i>	%	<i>n</i>	%		<i>n</i>	%	<i>n</i>	%		<i>n</i>	%	<i>n</i>	%	
	HH		HC			HH		HC			HH		HC		
computer human	65	48.5	58	49.6	0.967	44	49.4	31	51.7	0.921	21	46.7	27	47.4	0.897
	69	51.5	59	50.4		45	50.6	29	48.3		24	53.3	30	52.6	
	CH		CC			CH		CC			CH		CC		
	<i>n</i>	%	<i>n</i>	%		<i>n</i>	%	<i>n</i>	%		<i>n</i>	%	<i>n</i>	%	
computer human	61	45.5	66	56.4	0.111	46	51.7	39	65.0	0.149	15	33.3	27	47.4	0.220
	73	54.5	51	43.6		43	48.3	21	35.0		30	66.7	30	52.6	

In the analysis, no differences were observed regarding the composer or interpreter of the music (the human or the computer) and this lack of bias was observed for both professional musicians and non-musicians. The results are shown in Table 9. Musicians showed marginally higher values and a slight tendency to classify the piece CC as computer made. In general, all subjects failed to correctly identify the compositional source of the two musical samples.

Table 9. Sensitivity and specificity in the human-computer judgement.

	Musicians			Non-Musicians		
	Value	CI (95%)		Value	CI (95%)	
sensitivity (%)	57.05	48.76	65.33	41.18	31.14	51.22
specificity (%)	49.66	41.3	58.03	52.94	42.76	63.12

4. Discussion and Conclusions

In this paper we have presented a system that combines formal grammars and evolutionary algorithms to compose in both atonal and tonal music styles. We have described how it works internally and how, differently from many other approaches, this one does not require a pre-existing dataset of compositions. This property brings a set of important advantages:

- The product is innovative, since it is not based on imitation and has complete freedom to explore the space of search defined by the (more restrictive or looser) input rules, which act merely as a check that the new samples comply with the commission. Apart from the fitness and the evolutive mechanisms, the other essential part of the system that allows this free search is its implicit encoding based on formal grammars. They impose a hierarchical structure and favor behaviours like repetitions of musical units, satisfying some of the basic requirements of music composition.
- The syntax to write the genomes (and hence the music) is (a) highly expressive, since it allows the representation of any piece of music that can be expressed in the common music notation; (b) flexible, any musical sequence can be written in infinite forms; (c) compact, meaning that in spite of including all the compositional and performing information, it consumes between a half and a third less storage than the equivalent in MIDI format, between a third and a quarter of a corresponding MusicXML file and definitely less than any audio based format; and (d) robust, meaning that if a genome is altered in any way, not only it still produces a valid piece of music, but it also shares many elements in common with the original, as being a mutation of it.
- The system is affordable: (a) to set up, since there is no need to search and obtain samples from any external source; (b) regarding memory space, since there is no need to store or move large amounts of data to train; (c) while it is true that it requires the

intervention of an expert to input the rules and achieve convergence to a particular style, once it is done, the execution is not very computationally demanding. Using a single CPU thread in a current computer, both the atonal and the tonal systems roughly produce one genuine composition in their most complex style in 6.5 min; and it is possible to run in parallel as many of these tasks as wanted.

The system is set up at the beginning, if desired, by using just a few simple highly abstract parameters, such as amount of dissonance, repetitiveness, duration of the composition, etc., which are then translated to the genomes that undergo the evolving process. It neither requires any existing creative input from the user nor does it require an interactive interaction from them (i.e., interactive fitness function). For a more specific purpose, the system can be constrained by allowing people with musical knowledge to produce music in a particular style, in roughly ten minutes, through a set of global parameters (harmony, rhythm, instruments, etc.) that act as fitness and can still be considered highly abstract musical directions, nothing more than what would be given to an expert musician to compose.

However, normally the composition of music is regarded as a creative process where humans can express and evoke sensations. Could we ever consider this artificially-created music, actual music? The definition of music is a tricky question, even if only instrumental music is considered. For example, thinking of it as an organized set of sounds can be too broad an understanding, as it includes a collection of sounds that, while organized, are not music and, at the same time, it takes the risk of excluding the compositions made by Melomics, since until they are synthesized (or performed) they are just scores. Two properties that are usually required in the definition of music are either tonality—or the presence of certain musical features—and an appeal to esthetic properties. The first can be guaranteed by Melomics through the encoding and the fitness function. The second property is certainly more complex to be respected, since we are unable to ascribe any esthetic intention to the software, but if we move the focus from the composer to the listener, the fact that the composer is a machine is not as relevant. Among the different comments on Melomics music, Peter Russell's [53] positive judgement was interesting to us as he, who had no knowledge of the origin of the composition, does not express any doubt on the musical nature of the piece. This encouraged us to follow this understanding to assess whether the system generated actual music. The mentioned properties are not the only way to define music and, indeed, they have problems in capturing all and only what humans usually consider music. There are multiple possible definitions, each one with strengths and weaknesses and their discussion is more the domain of the philosophy of music [54]. We needed to take a more practical approach, instead of looking at Melomics's compositions from a more philosophical perspective, we considered the opinion of critics and of the general public. If they considered the end result as music, then it could be considered music.

The TT was designed to identify thought traces in computer processing and its interactive nature makes the adaptation to a musical version difficult. Nevertheless, the underlying principle of Turing's approach remains a valid inspiration for new tests that measure how close artificial music from human music is. In contrast with previous works, the experiment presented illustrates a controlled and rigorous methodology for a trial of this nature performed over a large sample of participants. The first question of the questionnaire was motivated to measure potential differences perceived from the original music sample composed by a musician and the other samples. In this sense, it is worth noting that the natural sounds sample was classified as music by 41.7% of the professional musicians. In contrast, both music samples were classified as music by most of the subjects (over 90%). With respect to the capability of eliciting mental images, it also endorsed the hypothesis of this experiment, as there was no significance when evaluating the musical samples, eliciting images in around 50% of the subjects. With respect to the natural sounds sample, this measure raised to 90% (both in the specific question and in the presence of terms in the descriptions). This is not surprising: natural sounds are associated with concrete and

recognizable physical sources, while music is produced by musical instruments and images arise in the form of memories not directly related to the perceived sounds, following a more abstract channel. The study of qualitative data confirmed this fact: most of the terms used in these descriptions fit the category “nature”, differing from those used to describe the musical samples, which do not point to any of the defined categories. These results highlight the difference between natural sounds and the presented music, which appears to generate a more complex and wider set of mental images, with independence of the musical training of the listener or who composed or interpreted the pieces. With respect to the evoked emotional states, one of the most revealing results was that natural sounds had a significant rate of 89.7% of descriptions assigned to the state “calm”, in contrast to the music recordings, with a maximum rate of 47.3% in this category, even though the music style presented was arguably calm. As in the case of mental images, all the musical pieces seemed to elicit a wider range of emotions, with independence of the listener, the composer or the interpreter. The second interesting result came from the study of valence in the descriptions. They turned out to be significantly positive ($p < 0.001$) when describing sounds of nature, while in the case of music, with no relevant differences among the groups of study, they also elicited unpleasant feelings. The final part of the test confirmed that the subjects were unable to distinguish the source of the composition. Even if it was done only with two different musical pieces, the sample was wide and the fact that about a half of it was made of professional musicians is an indicator of the robustness of the conclusions, which confirms the hypothesis, suggesting that computer compositions might be used as “true music”. This is, of course, a first result in this line of quality assessment. The model can be extended to different musical styles, given that the automatic music synthesis module (in that particular style) is good enough as considered by the subjects involved in the present experiment.

There are a lot of cases in daily situations that can benefit from music being produced automatically, especially those cases that are not the focus of composers. It can make music composition accessible to more people, even those with very little or no knowledge of the process, which could eventually lead to new styles. A particular application that we would like to develop more in the future is adaptive music. The potential of generative music that complies with human music allows the creation of tailored music (specific genres, structure, instrumentation, tempo, rhythm, etc.) considering personal preferences as well as particular needs or goals, such as responding in real time to the evolution of physiological signals. This can be used for example in therapy (pain relief, sleep disorders, stress, anxiety) or to assist during physical activity.

Supplementary Materials: The following are available online at <https://www.mdpi.com/2076-3417/11/9/4151/s1>, Audio S1: Atonal Development Iteration 0; Audio S2: Atonal Development Iteration 1; Audio S3: Atonal Development Iteration 2; Audio S4: Atonal Development Iteration 3; Audio S5: Tonal Development Iteration 4; Audio S6: Nokia tune mutated; Audio S7: Original DiscoWow2 theme; Audio S8: DiscoWow2 theme crossed with TV tune; Audio S9: Perceptive Study Audio Samples; Scores S10: Atonal pieces; Theme S11: Tonal piece.

Author Contributions: Conceptualization, D.D.A.-M. and F.J.V.; methodology, D.D.A.-M., F.J.V. and A.R.; software, D.D.A.-M.; validation, D.D.A.-M., F.J.V., A.R. and F.R.-R.; formal analysis, D.D.A.-M., F.J.V. and F.R.-R.; investigation, D.D.A.-M.; resources, F.J.V.; data curation, D.D.A.-M. and F.J.V.; writing—original draft preparation, D.D.A.-M.; writing—review and editing, D.D.A.-M. and F.J.V.; visualization, D.D.A.-M., F.J.V. and F.R.-R.; supervision, F.J.V.; project administration, D.D.A.-M. and F.J.V.; funding acquisition, D.D.A.-M. and F.J.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Spanish Ministry of Economy and Competitiveness under grant IPT-300000-2010-010, Spanish Ministry of Industry, Energy and Tourism under grant TSI-090302-2011-8 and Research programme of Junta de Andalucía under grant GENEX.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The relevant data generated in this study is contained within the article or provided as supplementary material.

Acknowledgments: The authors are grateful to the Museum for Interactive Music of Malaga (MIMMA) and the Conservatory of Málaga for providing the facilities to implement the experiments. Also to Carlos Sánchez-Quintana who collected data at MIMMA and Juan Carlos Moya and Nemesio García-Carril Puy for interesting discussions on the philosophical interpretations of Melomics.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dacey, J. Concepts of creativity: A history. In *Encyclopedia of Creativity*; Runco, M.A., Pritzker, S.R., Eds.; Elsevier: Amsterdam, The Netherlands, 1999; Volume 1, pp. 309–322.
2. Boden, M.A. Computer models of creativity. *AI Mag.* **2009**, *30*, 23. [\[CrossRef\]](#)
3. Pachet, F.; Cazaly, D. A taxonomy of musical genres. In Proceedings of the Content-Based Multimedia Information Access Conference (RIAO), Paris, France, 12–14 April 2000; pp. 1238–1245.
4. Fernández, J.D.; Vico, F. AI methods in algorithmic composition: A comprehensive survey. *J. Artif. Intell. Res.* **2013**, *48*, 513–582. [\[CrossRef\]](#)
5. Ariza, C. Two pioneering projects from the early history of computer-aided algorithmic composition. *Comput. Music. J.* **2011**, *35*, 40–56. [\[CrossRef\]](#)
6. Hiller, L.A., Jr.; Isaacson, L.M. Musical composition with a high speed digital computer. *J. Audio Eng. Soc.* **1957**, *6*, 154–160.
7. Ames, C. Automated composition in retrospect: 1956–1986. *Leonardo* **1987**, *20*, 169–185. [\[CrossRef\]](#)
8. Olson, H.F.; Belar, H. Aid to music composition employing a random probability system. *J. Acoust. Soc. Am.* **1961**, *33*, 1163–1170. [\[CrossRef\]](#)
9. Gill, S. A Technique for the Composition of Music in a Computer. *Comput. J.* **1963**, *6*, 129–133. [\[CrossRef\]](#)
10. Cope, D. Computer modeling of musical intelligence in EMI. *Comput. Music. J.* **1992**, *16*, 69–83. [\[CrossRef\]](#)
11. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Proceedings of the 27th International Conference on Neural Information Processing Systems, Montréal, Canada, 8–13 December 2014; pp. 2672–2680.
12. Eck, D.; Schmidhuber, J. Finding temporal structure in music: Blues improvisation with LSTM recurrent networks. In Proceedings of the 2002 12th IEEE Workshop on Neural Networks for Signal Processing, Martigny, Switzerland, 6 September 2002; pp. 747–756.
13. Colombo, F.; Seeholzer, A.; Gerstner, W. Deep artificial composer: A creative neural network model for automated melody generation. In Proceedings of the International Conference on Evolutionary and Biologically Inspired Music and Art, Amsterdam, The Netherlands, 19–21 April 2017; pp. 81–96.
14. Gero, J.S. Computational models of innovative and creative design processes. *Technol. Forecast. Soc. Chang.* **2000**, *64*, 183–196. [\[CrossRef\]](#)
15. Pearce, M.; Wiggins, G. Towards a framework for the evaluation of machine. In Proceedings of the AISB’01 Symposium on Artificial Intelligence and Creativity in the Arts and Sciences, York, UK, 21–24 March 2001; pp. 22–32.
16. Ritchie, G. Some empirical criteria for attributing creativity to a computer program. *Minds Mach.* **2007**, *17*, 67–99. [\[CrossRef\]](#)
17. Biles, J. Genjam: A genetic algorithm for generating jazz solos. In Proceedings of the 1994 International Computer Music Conference, Aarhus, Denmark, 12–17 September 1994; Volume 94, pp. 131–137.
18. Moroni, A.; Manzolli, J.; von Zuben, F.; Gudwin, R. Vox populi: An interactive evolutionary system for algorithmic music composition. *Leonardo Music. J.* **2000**, *10*, 49–54. [\[CrossRef\]](#)
19. Nam, Y.W.; Kim, Y.H. Automatic jazz melody composition through a learning-based genetic algorithm. In Proceedings of the International Conference on Computational Intelligence in Music, Sound, Art and Design, Leipzig, Germany, 24–26 April 2019; pp. 217–233.
20. Lindemann, A.; Lindemann, E. Musical organisms. In Proceedings of the International Conference on Computational Intelligence in Music, Sound, Art and Design, Parma, Italy, 4–6 April 2018; pp. 128–144.
21. Gilbert, É.; Conklin, D. A probabilistic context-free grammar for melodic reduction. In Proceedings of the International Workshop on Artificial Intelligence and Music, 20th International Joint Conference on Artificial Intelligence, Hyderabad, India, 6–12 January 2007; pp. 83–94.
22. Abdallah, S.; Gold, N.; Marsden, A. Analysing symbolic music with probabilistic grammars. In *Computational Music Analysis*; Meredith, D., Ed.; Springer: Cham, Switzerland, 2016; pp. 157–189.
23. Rohrmeier, M. Towards a generative syntax of tonal harmony. *J. Math. Music.* **2011**, *5*, 35–53. [\[CrossRef\]](#)
24. Lindenmayer, A. Mathematical models for cellular interaction. *J. Theor. Biol.* **1968**, *18*, 280–315. [\[CrossRef\]](#)
25. McCormack, J. A developmental model for generative media. In *European Conference on Artificial Life*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 88–97.

26. Rodrigues, A.; Costa, E.; Cardoso, A.; Machado, P.; Cruz, T. Evolving l-systems with musical notes. In Proceedings of the International Conference on Computational Intelligence in Music, Sound, Art and Design, Porto, Portugal, 30 March–1 April 2016; pp. 186–201.
27. Turing, A.M. Computing machinery and intelligence. *Mind* **1950**, *59*, 433–460. [CrossRef]
28. Harnad, S. Minds, machines and Turing: The Indistinguishability of Indistinguishables. *J. Logic Lang. Inf.* **2000**, *9*, 425–445. [CrossRef]
29. Bringsjord, S.; Bello, P.; Ferrucci, D. Creativity, the Turing test, and the (better) Lovelace test. *Minds Mach.* **2001**, *11*, 3–27. [CrossRef]
30. Ariza, C. The interrogator as critic: The Turing test and the evaluation of generative music systems. *Comput. Music J.* **2009**, *33*, 48–70. [CrossRef]
31. Papadopoulos, G.; Wiggins, G. AI methods for algorithmic composition: A survey, a critical view and future prospects. In Proceedings of the AISB Symposium on Musical Creativity, Edinburgh, Scotland, 6–9 April 1999; pp. 110–117.
32. Delgado, M.; Fajardo, W.; Molina-Solana, M. Inmamusys: Intelligent multiagent music system. *Expert Syst. Appl.* **2009**, *36*, 4574–4580. [CrossRef]
33. Monteith, K.; Martinez, T.; Ventura, D. Automatic generation of music for inducing emotive response. In Proceedings of the International Conference on Computational Creativity, Lisbon, Portugal, 7–9 January 2010; pp. 140–149.
34. Roig, C.; Tardón, L.J.; Barbancho, I.; Barbancho, A.M. Automatic melody composition based on a probabilistic model of music style and harmonic rules. *Knowl. Based Syst.* **2014**, *71*, 419–434. [CrossRef]
35. Williams, D.; Kirke, A.; Miranda, E.R.; Roesch, E.; Daly, I.; Nasuto, S. Investigating affect in algorithmic composition systems. *Psychol. Music.* **2015**, *43*, 831–854. [CrossRef]
36. Juslin, P.N.; Sloboda, J. *Handbook of Music and Emotion: Theory, Research, Applications*; Oxford University Press: Oxford, UK, 2011; pp. 187–344.
37. Meyer, L.B. *Emotion and Meaning in Music*; University of Chicago Press: Chicago, IL, USA, 1956.
38. Koelsch, S. Brain correlates of music-evoked emotions. *Nat. Rev. Neurosci.* **2014**, *15*, 170–180. [CrossRef] [PubMed]
39. Koelsch, S. Music-evoked emotions: Principles, brain correlates, and implications for therapy. *Ann. N. Y. Acad. Sci.* **2015**, *1337*, 193–201. [CrossRef] [PubMed]
40. Ohno, S. Repetition as the essence of life on this earth: Music and genes. In *Modern Trends in Human Leukemia VII*; Neth, R., Gallo, R.C., Greaves, M.F., Kabisch, H., Eds.; Springer: Berlin/Heidelberg, Germany, 1987; pp. 511–519.
41. Bent, I.D.; Pople, A. Analysis. In *Grove Music Online*; Oxford University Press: Oxford, UK, 2010. [CrossRef]
42. McKay, C. Automatic Music Classification with jMIR. Ph.D. Thesis, McGill University, Montréal, QC, Canada, 2010.
43. McKay, C.; Fujinaga, I. The Bodhidharma system and the results of the MIREX 2005 symbolic genre classification contest. In Proceedings of the ISMIR 2005, 6th International Conference on Music Information Retrieval, London, UK, 11–15 September 2005.
44. The Echo Nest. Available online: <https://github.com/echonest> (accessed on 30 April 2021).
45. The Sounds of Nature Collection. 2010. Available online: <https://gumroad.com/l/nature> (accessed on 30 April 2021).
46. Animals/Insects. Available online: http://www.freesfx.co.uk/soundeffects/animals_insects/ (accessed on 30 April 2021).
47. Steinbeis, N.; Koelsch, S. Understanding the intentions behind man-made products elicits neural activity in areas dedicated to mental state attribution. *Cereb. Cortex* **2009**, *19*, 619–623. [CrossRef]
48. Moffat, D.C.; Kelly, M. An investigation into people’s bias against computational creativity in music composition. In Proceedings of the 3rd international joint workshop on computational creativity (ECAI06 Workshop), Riva del Garda, Italy, 28–29 August 2006; pp. 1–8.
49. Castellaro, M. El concepto de representación mental como fundamento epistemológico de la psicología. *Límite* **2011**, *24*, 55–68.
50. Díaz, J.L.; Enrique, F. La estructura de la emoción humana: Un modelo cromático del sistema afectivo. *Salud Ment.* **2001**, *24*, 20–35.
51. Plutchik, R. The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *Am. Sci.* **2001**, *89*, 344–350. [CrossRef]
52. Russell, J.A. A circumplex model of affect. *J. Personal. Soc. Psychol.* **1980**, *39*, 1161–1178. [CrossRef]
53. Ball, P. Artificial music: The computers that create melodies. *BBC* **2014**. Available online: <https://www.bbc.com/future/article/20140808-music-like-never-heard-before> (accessed on 30 April 2021).
54. Kania, A. The philosophy of music. *Stanf. Encycl. Philos. Arch.* **2007**. Available online: <https://plato.stanford.edu/entries/music/> (accessed on 30 April 2021).