applied sciences

MDPI

Article

Technical, Musical, and Legal Aspects of an AI-Aided Algorithmic Music Production System

Joanna Kwiecień [1], Paweł Skrzyński [2], Wojciech Chmiel [1], Andrzej Dąbrowski [3], Bartłomiej Szadkowski [3] and Marek Pluta [4,*]

[1] Department of Automatic Control and Robotics, AGH University of Krakow, Av. Mickiewicza 30, 30-059 Krakow, Poland; kwiecien@agh.edu.pl (J.K.); wch@agh.edu.pl (W.C.)
[2] Department of Applied Computer Science, AGH University of Krakow, Av. Mickiewicza 30, 30-059 Krakow, Poland; skrzynia@agh.edu.pl
[3] Independent Digital Sp. z o.o., 00-344 Warsaw, Poland; kontakt@andrzejdabrowski.pl (A.D.); bartlomiej@kancelariaszadkowski.com (B.S.)
[4] Department of Mechanics and Vibroacoustics, AGH University of Krakow, Av. Mickiewicza 30, 30-059 Krakow, Poland
* Correspondence: pluta@agh.edu.pl; Tel.: +48-12-617-3416

Abstract: Even though algorithmic composition might be considered a centuries-old concept, it has been gaining particular momentum since the introduction of computer-based techniques. The development of artificial intelligence (AI) methods, culminating in the latest achievements of deep learning techniques, has provided tools to automatically compose and even produce music. This paper discusses various aspects of the entire process within a context of designing a system able to automatically generate a score and recordings belonging to selected musical genres. It begins with the idea and design overview, followed by considerations regarding the algorithmic formulation of selected musical rules and principles. The system implements a hybrid approach, combining conventional, i.e., stochastic or rule-based, and AI elements. The latter are applied to facilitate the generation of selected layers of composition and to constitute a classifier with a task of evaluating the generated recordings. Selected stages of music generation are discussed, for example how motifs are processed into phrases and how phrases are used in the context of a whole song. To validate the system operation results, an evaluation of the quality of the produced music recordings was conducted, including a test with a group of listeners. The analysis also touches upon some legal aspects related to the creation of algorithmic compositions.

Keywords: music production; artificial intelligence; algorithmic composition; music classification; automatic music creation

Citation: Kwiecień, J.; Skrzyński, P.; Chmiel, W.; Dąbrowski, A.; Szadkowski, B.; Pluta, M. Technical, Musical, and Legal Aspects of an AI-Aided Algorithmic Music Production System. *Appl. Sci.* **2024**, *14*, 3541. https://doi.org/10.3390/app14093541

Academic Editor: Lamberto Tronchin

Received: 27 February 2024
Revised: 16 April 2024
Accepted: 17 April 2024
Published: 23 April 2024

1. Introduction

For a long time, music has been a domain of human artists, but this state is subtly changing. New sophisticated tools allow creators to concentrate on ideas or concepts, while the most arduous tasks can be taken away and carried out automatically. A leap has been made with the introduction of streaming services. Combined with advances in audio engineering and artificial intelligence (AI) algorithms, it provides musicians with new opportunities. They are given easy to use, broadly available tools, matched by means to quickly publish new music worldwide and gather almost immediate feedback. These advances benefit not only creators but also consumers—more people can be involved on both sides. With global availability, each music genre finds its niche. Therefore, a situation where convenient and advanced tools for creators are accompanied by streaming services and popularity of streamed music consumption encourages musical experiments. Among the most interesting are the experiments in automatic creation of music. Past approaches were limited predominantly to algorithmic composition, but the full process of music production has only recently become possible to consider. And it brings numerous issues.

Music production is a multi-stage process that leads to the creation of music, usually in the form of an audio recording. It involves composition, sound design, arrangement, mixing, mastering, and, last but not least, a critical evaluation of the result. As far as it is known, no solution in automatic production has been proposed yet to evaluate generated recordings by building a complex, realistic generator-critic system.

In the proposed approach, a unit referred to as the Generator is responsible for all the stages from composition to mastering. It creates music and generates an audio file, given a set of simple requirements from a user. In a utilitarian use case, a user might want to specify a genre of music and some of its general features, such as mood, to make it suitable for a particular purpose. Various algorithms and heuristics implemented in Generator, including AI, substantiate these requirements into a ready-to-use music file. Such an automatic process may not always lead to satisfactory results. Therefore, it has been coupled with an AI supervisor, with a sole purpose of critical evaluation of the output of the Generator. Depending on the evaluation, the generated music is either accepted and presented to a user or the process of generation is repeated.

Basically, there are two kinds of classification techniques in the literature: non-supervised and supervised. Some approaches tend to grouping of music data in a non-supervised way, so that by using various similarity measures, a classification will arise from the data themselves. Here, the supervised approach to recording classification has been used. Taking into account ambiguity, recordings could be associated with a variety of general labels. Thus, this work considers that classification involves only two labels of generated recordings: 'good' and 'incorrect'. Note that various algorithms could be used to solve a wide range of classification problems, e.g., logistic regression, the naive Bayes classifier, support vector machine (SVM), k-nearest neighbors (kNN), decision trees, or neural networks [1,2]. In the presented solution, a neural network is used as a classifier (referred to as the Critic).

1.1. Contribution

The goal of this paper is to present selected topics related to algorithmic music production with AI methods and to describe some technical solutions in our actual, complex, realistic, and robust generator-critic system of automatic music generation that produces not only melodies but also whole, multi-instrumental music recordings that are acceptable to a listener within selected music genres. Most contemporary creative AI mechanisms are "fed" a large amount of musical content, which raises legal questions, making some organizations or businesses prohibit works of such origin. In this regard, the approach presented in this study is novel by selecting AI methods without such legal defects while still being able to produce complex works. To validate the results of our system and to meet the requirements of people interested in music (dance, electronic, and relaxation), we evaluated the quality of produced music recordings by conducting tests with a selected group of listeners using the method described in [3]. Algorithmic music production is a real technical and legal challenge. Hence, another contribution of this article is signaling that the legal aspects must be taken into account when creating "artificial" music.

1.2. Related Work

The system discussed in this paper is related to two well known concepts, i.e., functional, or background music, and algorithmic composition. Both were conceived centuries before the advent of digital computers. A characteristic of functional music is that it has a secondary, accompanying role during some other activity. It has to comply with various requirements regarding its quality, and therefore it may be well suited for some algorithmic approaches that help to achieve predictable results [4–6]. This kind of music, where artistic value is secondary to utilitarian value and adaptation to externally imposed requirements, is a suitable target for the presented system. A good example may be music created for the purpose of relaxation.

Algorithmic composition is a technique that easily dates back to at least the early polyphonic music and rules of counterpoint [7]. In the 20th century, it helped to develop the dodecaphony and serial techniques [8], but for the most part, it was considered a composer's tool. It required an introduction of digital computers to code and apply selected rules so that entire works could be composed in a fully automated manner. First among notable examples was a string quartet 'Illiac Suite' by Hiller and Isaacson [9]. The principle applied was based on the counterpoint and dodecaphonic rules controlled by probabilistic techniques.

After the 'Illiac Suite', various methods and techniques have been applied to music generation. In [10], seven categories were discussed, such as Markov chains, formal grammars, rule-based systems, neural networks (with deep learning), evolutionary algorithms, chaos similarity, and agent-based systems.

AI methods have greatly expanded in recent years. Finding the needed method has turned out to be easier, necessary, and sufficient for the generation and classification of music data, which is a complex task. As reported in [11], several AI techniques and algorithms have been proposed for solving the problem.

Evolutionary algorithms have been applied in various areas of music composition. Many directions of their application for composing music can be found, e.g., making variations of an existing composition or motif, considering a melody composition with or without rhythm generation, and so on. In [12], a genetic algorithm that evolved a four-part musical composition melodically, harmonically, and rhythmically was presented. In [13], the use of evolutionary computation to create melodic line harmonization was proposed. In turn, a survey by Briot et al. [14] specifically focused on an analysis of using deep learning for music generation. Recently, many studies have explored deep neural networks for creating music and obtaining its proper quality [15]. In the domain of algorithmic music composition, promising results have been described using deep recurrent neural networks, mainly a specific recurrent neural network known as the long short-term memory neural network or simply LSTM. For example, in [16], LSTM for polyphonic music prediction was proposed. In [17], a method for automatically developing music based on music segments from existing music and LSTM for tinnitus music therapy was described. Moreover, some studies have observed good results when examining the transformers to create music compositions. For example, Google used the transformer for music generation [18]. In [19], a music generation model that combines a transformer with generative adversarial networks (GANs) was proposed. Neves et al., in [20], proposed a transformer-GAN model to create symbolic music conditioned by sentiment. In [21], Jin et al. proposed a scheme based on combining the transformer and generative pre-training model (GPT) to generate multi-track music including tracks of piano, guitar, and drum. A good review of the available works about AI methods and solutions is given in [22].

A number of papers have discussed various classification problems. One of the main research topics is genre classification, which is assigning music items into different predefined genres. Automatic music genre classification as a pattern recognition problem has been described in [23]. To classify music according to its genre, a set of features (timbral texture features, beat-related features, and pitch-related features) that represent music signals and three types of classifiers (Gaussian Classifier, Gaussian Mixture Models, and kNN) have been discussed. In [24], the use of both audio and symbolic descriptors for genre classification by SVM was proposed. Moreover, there have been some trials using deep learning in the field of music information retrieval. For example, in [25], a recurrent neural network with a channel attention mechanism for music feature classification was proposed. Also, ref. [26] studied music style classification based on deep learning methods and described a music classification module based on a one-dimensional convolution of recurring neural networks. In [27], by using a combination of convolutional neural networks and variants of recurrent neural networks (LSTM, bidirectional LSTM, gated recurrent units, and bidirectional gated recurrent units), a music classification task was

performed. Moreover, the subject of intensive research in the literature has been artist classification [28] and mood classification [29,30].

It is also important to mention the works aimed at evaluating the quality of music. For example, ref. [31] is devoted to measuring and evaluating musical excerpts of classical piano music generated by deep learning models based on the results of the blind test conducted on a group of musicians and non-musicians. In turn, in [32], an evaluation method for music (generated by deep learning) was proposed as a combination of mathematical statistics and music theory evaluation. However, this approach is not possible in our case due to the complexity of composed music. Claimed objectivity of such methods is based on arbitrary selection of rules, which in natural music changes over history and varies between styles. They may evaluate similarity between fragments of melody and rhythm [32] but not their objective quality due to a lack of common, generally agreed upon criteria for the aforesaid evaluation. Strict criteria can only be applied to narrow, specific cases, such as vocal melody in Renaissance counterpoint. Moreover, most studies regarding AI-generated music do not actually deal with music but with one or two of its elements only. It is usually a melody, i.e., order of pitches (scalars), and sometimes a rhythm, i.e., sequence of time intervals between note onsets or rests, expressed as multiples of a unit duration (scalars). Aspects such as harmony (chords and chord progressions), musical form (large-scale structure), choice and handling of instruments, dynamics, articulation, and tempo (and its variability), and timbre-related mixing and mastering are either ignored or considered only marginally. The evaluation of individual elements of music does not provide reliable results: a human listener can evaluate the same melody differently if it is played in a different tempo, with a different harmonic background, or with a different instrument. Therefore, in case of the automatic production of complete musical recordings combining all the aforesaid elements, at present, the only viable solution is to use a panel of listeners. In the future, it can be replaced with a deep learning model, but it will require arduous training with real listeners—at the moment, it works only in a very narrow set of cases [31].

2. Materials and Methods

The aim of this section is to explain the process that is applied in the presented system to produce recordings as well as the practical use of AI methods for an algorithmic composition.

2.1. System Overview

As a first step, our system generates recordings belonging to one of three possible genres, i.e., relaxation, electronic, and dance music. Then, a neural network (trained with generated labeled data) verifies the result through classification of the recordings. The dataset for each genre consists of 2000 audio recordings generated by our system and labeled by experts. We consider only a case in which we build a classifier separately for each genre. The classifier returns the score of a recording within a genre (1 for "good" or 0 for "incorrect"). Therefore, we present an overview of the essential tasks in our automatic system for simultaneous generation and classification of recordings.

The Generator produces music in the form of an audio recording, indirectly controlled by a user through a set of adjustable features. Internally, the generated music goes through several stages of representation, including a musical score and sequencer data. The latter controls a synthesizer that converts symbolic data into an audio signal. The process of generation itself is carried out without the user's interaction in a time much shorter than the duration of generated music. It is possible to generate several new music files while still reproducing a previous one. There is enough time for the Critic to perform an evaluation of an output file, and—in case of a negative result—to send a request for a new one to the Generator. The evaluation itself is based not on the musical score data but on the parameters of output recordings to take into account the results of synthesis, mixing, and mastering.

One of the assumed requirements is that the Critic uses only a set of parameters describing generated recordings. Clearly, the key objective of the created classifier is

to ensure the best chance of success for the automatic music production. Concerning a classification process, the stage of feature extraction is necessary because the audio signals contain redundant or irrelevant information. We obtain information on the usefulness of generated recordings in the form of an assigned label as a result of using our Critic. In essence, it indicates how well the Generator produces recordings.

2.2. The Generator

Music organizes sound structures in time, although the actual forms these structures may assume vary depending on the stage of the creative process. The process starts with a concept that evolves into better defined, yet still abstract ideas. The ideas are transformed into a symbolic score, which is definite but may not express all the features of the previous stage. These are supplemented through an interpretation. Next, a performance produces a sonic representation that may be recorded, processed, and stored as an audio signal to be later reproduced, listened to, and perceived. A similar model of data flow has been adapted for the purpose of automatic music production (Figure 1).

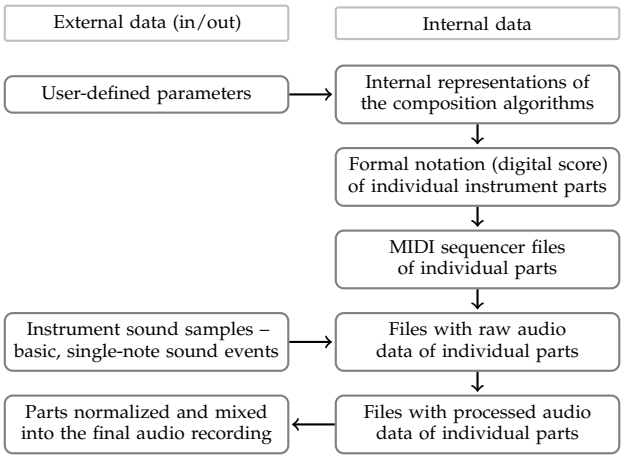

Figure 1. Data in subsequent phases of automatic music production.

User-provided data represent a concept. The Generator uses data structures such as arrays (single type), lists (multiple types), maps (key-value), or rule sets to internally represent higher-level abstract musical structures throughout composition algorithms. The algorithms produce digital scores for individual tracks based on different principles, with elements of interpretation coded as performance marks (dynamics and articulation). Scores are converted to sequencer control data (MIDI—Musical Instrument Digital Interface) and supplemented with interpretation elements (pitch tuning). Sequencer control of a non-real-time sampler through MIDI is an equivalent of a performance and a recording (an output is directed into an audio file). Audio tracks are processed and mixed to undergo a mastering process.

The process applied by the Generator to produce a musical work is presented in Figure 2. It may be assumed that in automatic production, a default approach would be to allow a user to define a general concept only. Therefore, in the first stage, a user provides a set of scalar parameters, as described in Table 1, controlling basic features of a single work. These characteristics have multiple relations with a larger set of internal algorithm parameters and affect various phases of the production process. At some point, there will emerge a need to allow a larger degree of control; thus, a set can be expanded with more specific parameters.

As an example, the two most important aspects controlled by the mood parameter are the scale and chord type selection probability and the prevalence of a particular movement type in rhythmic motifs. Some musical scales as well as some chord types are considered happy, whereas some are considered sad; therefore, the parameter controls the probability of occurrence of either type in a recording (recordings utilize more than one scale or chord

type). With regards to rhythm, setting a higher mood value (happy) allows more frequent use of shorter rhythmic values and characteristic groups, such as fast dotted rhythms. One might also consider tempo as strongly related to mood, which is true, but in a production system, the tempo is a key parameter and has to be controlled directly, so it has been left independent from the mood parameter.

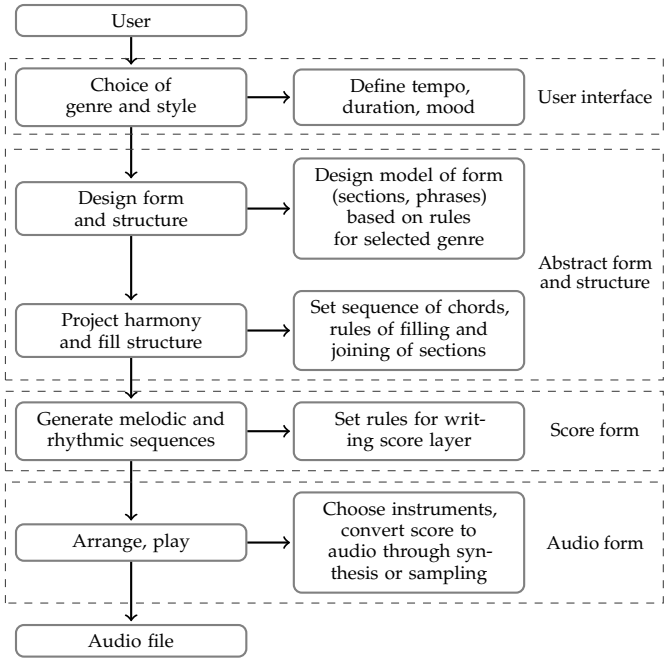

Figure 2. The model of the music generation process.

Table 1. User-controlled parameters defining a single musical work.

Data	Type	Comment
Genre	Enumeration	Single choice from a list
Duration	Integer	[s]
Tempo	Integer	[BPM]
Mood	Floating point	Range from *sad* to *happy*
Oddity	Floating point	Range from *normal* to *odd*
Tuning	Floating point	Fundamental frequency of A4 [Hz]

Another example is the oddity parameter. It has no common meaning in music, but in our system, it has been designed as a way to allow the generation process to differ more or less from the initially defined standard, common sense settings. If set to high values, it introduces more variety at the cost of a possibly not very well received musical solution. In particular, it controls the probability of non-typical harmonic sequences and allows the introduction of some non-typical scales. For the generation of motifs, it allows the use of some larger intervals more frequently, which can make melodies more interesting but also less easy to memorize. Also, in motif generation, a higher oddity introduces more motifs with less-ordered, irregular rhythmic sequences and flattens the initial probability distribution of note durations allowing a more frequent choice of extreme durations. In drum patterns. the oddity controls the amount of mutation applied to the initial base patterns.

The second stage is an analogue to the process of composition, where a series of operations eventually produces a formal musical score. A musical work may be seen as a hierarchic, multi-layer structure. Its design requires the interaction of multiple algorithms with different internal data representations.

At the top level, an overall structure of a composition is determined by a musical form. Automatic production may be based upon a set of common forms typical of particular genres and a parameter-guided series of choices. Choices are made based on genre-dependent probabilities and include the relation between harmony and form, i.e., the number of chords in a sequence, the size of chord areas compared to a phrase length, and the repeatability. Other decisions concern lengths of internal parts, which are related to a type of form as well. All of these decisions are made with regards to the total recording duration and the tempo, both of which are fixed and given by a user. Specific tempo–duration combinations can accommodate specific types of forms, which is another rule-based decision. Finally, instruments or tracks (such as ambient) are chosen, including a single or a pair of lead instruments—again, this choice is guided by the genre.

A form needs to be substantiated into a specific structure. This purpose may be served by the concept of patterns and order adapted from music trackers and groove boxes [33]. Thus, the structure of a composition is represented by a two-dimensional plan consisting of a set of instrument-assigned tracks, each with its own sequence of sections. A section represents a constant-length segment of a track, usually a few measures. It is common for musical material to be reused throughout a composition. Therefore, each track has its own set of unique sections to pick from, and the plan stores a sequence of section IDs, arranged in any order within tracks, including repetitions or omissions (Figure 3). It is here that particular forms, symmetric or asymmetric, such as a reprise form, rondo, variations, chorus–verse, and others, are substantiated into a layout, such as in Figure 3, considering internal sub-parts, use of instruments, breaks, etc. These choices are rule-based, with different rules for particular genres. Elements of fuzzy logic are used to guide probability-based decisions.

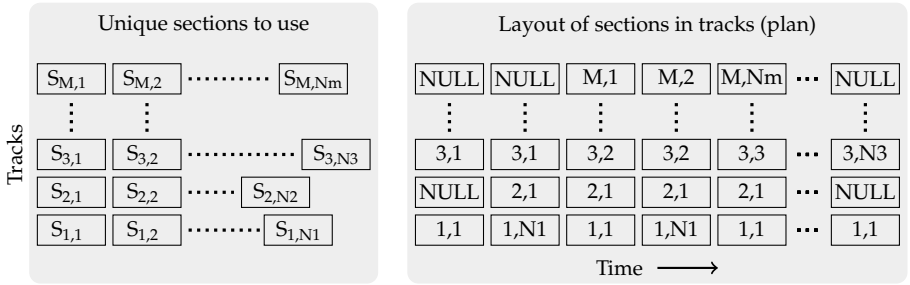

Figure 3. The structure of a composition with M tracks, each one using N_i unique sections $S_{i,j}$ (i—track index, j—unique section index within a track-assigned section list), and an example of a layout.

Before sections can be populated with rhythms and melodies, a layer of harmony needs to be considered. It imposes certain restrictions on the plan; hence, both need to be designed in parallel. The harmony is represented by a sequence of root pitch distances, measured in semitones in relation to a pitch assumed as an initial tonic. In many genres, such a sequence may be looped, and its length may vary between generated works. One section of a track in a plan of composition is assigned to a single step of the harmonic sequence. Longer sequences may be paired with shorter sections. Looped sequences are reflected in particular patterns within the plan. The type of harmonic sequence is related to the selected genre. The simplest one is used in relaxation and electronic genres only, though less frequently in the latter. It is based on the concept of minimal music and on the repetition of a single chord rooted in a single scale. A bit more complex sequence type consists of two alternating chords in one of several root relations, such as the dominant, subdominant, mediant, or next-step. This type can be used in all genres. The third type of sequence consists of a longer set of chords with a root and chord type order either designed freely, as in an alternating type, or chosen from a long list of predefined sequences, dedicated for particular genres. Chord types in a sequence are affected by the mood value.

Sections contain actual score data generated on the basis of various principles, depending on the instrument and its role in the composition. These data are symbolic, with pitch, duration, articulation, dynamics, etc., represented by sequences of characters. Each section is therefore stored in a single string type. It may be transposed or subjected to other musical transformations by addition of control commands for the score interpreter.

Sections of bass lines and accompanying tracks, such as chords or pads, are designed in analogy to accompanying parts in music workstations. A large set of predefined sequences and structures, typical for either rhythmic chords or pads, are stored for each genre. Sequences are fixed, but structures have elements, such as an order of selected pitches or an order of selected rhythmic values, that are decided during the generation process, thus allowing for a degree of variability.

Generator makes use of three AI-related techniques, as shown in Table 2, namely, the fuzzy logic (FL), genetic algorithm (GA), and rule-based systems (RBS). FL is the one used most widely on several stages as a means to soften edges between possible algorithm outcomes or to mix features. A notable example of FL usage is a gradual change in the selection probability of the mood and oddity parameter interpretation. Another interesting use of FL is in the lead melody design, where it guides a melody inversion in a proximity of the end-of-range for a particular instrument.

Table 2. Generator-side use of AI-related techniques: FL—fuzzy logic, GA—genetic algorithm, and RBS—rule-based system.

Task	FL	GA	RBS
Form design	+		
Harmonic progression generation	+		
Lead motifs generation	+		+
Lead phrases design		+	+
Drum patterns design		+	+
Bass lines generation	+		
Accompanying tracks design	+		+
Applying effects and mixing			+

Two remaining techniques are applied to produce sections of score for the instrument parts—often in combination with each other or with FL. RBS is applied to implement expert knowledge, such as rules of counterpoint, or some aesthetic recommendations. The RBS has a prominent role in the lead melody and phrase design. It uses counterpoint rules to evaluate various aspects of a generated phrase, such as a single-climax, no-excess-repetitions, avoiding particular interval sequences, etc. This evaluation is further used in the GA. In the final stages of production, the RBS is applied to control track-dependent settings in the mixing process. These are only a few examples, but RBS, being one of the most robust of the AI-related techniques, is used extensively throughout the entire system.

GA plays a key role in the generation of the most demanding parts—lead phrases and complex drum patterns. Lead phrases are designed from a limited set of initially prepared motifs to ensure consistency of the generated melodies. Motifs are constructed using rules controlling interval directions, jumps, repetitions, and a variability of rhythm. Selected motifs are transformed and combined in various ways to form a phrase, and the GA is used to evolve initial parameters of algorithms utilized within this process to finally obtain melodies best fulfilling good melody criteria according to counterpoint rules guided by the RBS. Another GA is applied in the drum pattern design to create new patterns from predefined ones with crossovers and mutations. Drum patterns are represented by binary arrays of instrument hits and skips, and the GA treats these arrays as specimens directly. This cannot be applied to lead phrases, which have a hierarchical, multi-layer structure and should not be modified directly. Here, a specimen represents a set of parameters for other algorithms that leads to the production of a score.

The transition from a score to an audio form is performed by the sequencer and the sampler. The former uses the MIDI data obtained directly from the score, whereas the

latter uses the prior genre-based selection of single-note instrument sound samples. It is not necessary to reproduce generated music in real time. Therefore, the sampler renders audio data directly to a file and can work faster than real-time. Depending on the computer hardware, at least several works can be produced within the time duration of a single work.

Selected sound effects are applied to audio files produced by the sampler, partially as a means to introduce final expressive touches. Effects are applied on the basis of the genre and the particular track and include an amplitude modulation (fluctuation or tremolo), low-pass and high-pass filters, the wah-wah effect, a bit crusher, and a dynamic compressor. LFO-controlled effects, such as an amplitude modulation or the wah-wah, can be synchronized to a rhythm. Processed files are mixed into a final recording. Mixing applies predefined signal level relations and panorama settings, with adjustments depending on, e.g., an actual track list. The settings for both the effects and mixing are determined using the knowledge-based RBS.

The system has been implemented as a set of Python scripts, with the main script executed either directly by a human operator or in a batch processing mode by an automation script. The Generator uses the following Python libraries—*os*, *random*, *time*, *math*, and *Mido*—for handling MIDI data [34] and *pyo* for audio signal processing [35], such as filtering, applying dynamic compression, envelopes, etc. No additional API is used throughout the generation process. All the algorithms are programmed directly in Python.

According to Figures 1 and 2, an abstract, internal representation of music is translated into a formal musical notation during the stage of track data generation. The notation uses the Lilypond score format [36], which is suitable for automatic composition. Lilypond is a music engraving program that works in a manner similar to TeX. It interprets music written as text files with logical, comprehensive syntax and produces a digital score in a graphical form as well as MIDI files for a sequencer.

MIDI files are fed into a MIDI sequencer that controls a sound synthesizer. The system requires both tools to work not in real-time, which would have been a default behavior, but to produce an audio signal as fast as possible, without an unnecessary wait time. This requirement severely limits available options. The system utilizes Fluidsynth [37], which produces audio files using sound samples of musical instruments stored in the SoundFont format [38]. The generated audio files contain separate tracks for the whole duration of the generated song.

Tracks are processed using the pyo library to add desired sound effects, some of which are tempo-synchronized, before the final mixing. Additional effects, such as reverb, are added using the SoX program [39]. The same SoX program is used to mix all the tracks into a single, stereo audio file.

An instrument selection is carried out according to specifications described in a user-editable data file. The file contains a list of assignments: a track name—a SoundFont file—an instrument preset number. Each track name has to appear in the data file at least once, meaning that it has to have at least one instrument assigned. It can, however, have multiple assignments, and in such a case, the Generator will pick one for a particular song. The same SoundFont and even the same instrument can be used in more than one type of track if the operator of the system decides so. There are currently eleven tracks, including two leads, various accompaniment tracks, and an ambient sound track.

The proposed hybrid approach has an advantage of greater flexibility over an approach based entirely on a single AI method. The use of various techniques on subsequent stages of music production, fit for particular tasks, allows the production of different genres of music and the control of details of the produced music in a meaningful manner, which is a crucial requirement for practical applications. Another advantage is the efficiency and the ability to run independently, without the need for a connection to the cloud. Finally, the proposed method is stable. It always produces an output that displays musical features and may be considered at least barely tolerable. Since not all produced works are of satisfactory quality, the Critic is applied to select only high-quality ones.

2.3. The Critic

One of the stages of algorithmic music production is the presence of the classification process. Popular classification methods are neural networks. In the proposed system, the multilayer feedforward neural network is used as a classifier. Between neurons in adjacent layers there are connections with associated weights that are adjusted during the neural network training procedure. In [40], an extensive survey on neural networks is presented.

Building a good method for classifying music with neural networks is not trivial. A good model should be able to distinguish different labels of generated recordings. Training and testing of the neural network are the main steps in the classification process. This process requires a data set that should be of sufficient size to guarantee proper learning. It becomes more difficult as the dimensions of the feature space increase. Therefore, a correct set of features is necessary for classification because good features separate classes and allow proper grouping. Of course, many features can be brought out, e.g., spectral, temporal, based on pitch, etc. Notably, the expert assessment of generated recordings should be taken into account. Therefore, among the main stages of the classification process, one can find as follows:

- creating, training, and testing sets,
- building, compiling, and training the neural network,
- evaluating the neural network on the test set.

At the beginning, we need to prepare a set of recordings from which the appropriate features will be extracted. Moreover, the expert assessment of the generated recordings is known. As it is known, music items can be represented with contextual information derived from available metadata (not encoded in the audio signal) or with an information extracted from audio content itself [41,42]. Therefore, in many studies, the procedure of music record classification involves the use of audio data-based musical content, symbolic data-based musical content, or hybrid information. Choosing the correct representation of the data is a key issue in the classification problem. Classical approaches (for example SVM, k-NN) use audio features such as mel-frequency cepstral coefficients (MFCCs) as input to a classifier. In turn, some deep learning methods (e.g., convolutional neural networks) use visual representations of the audio signal in the form of spectrograms. Specifically, the mel spectrogram can be taken as a visual representation of such a signal.

Therefore, one can use a set of selected spectral, time-domain, rhythm, and tonal descriptors. The following features to be extracted from the generated music and used in the classification process have been determined:

- low level descriptors, e.g.:
  - 13 first mel-frequency cepstral coefficients (MFCCs),
  - dissonance,
  - dynamic complexity,
  - pitch salience,
  - spectral complexity (Shannon entropy of a spectrum),
  - spectral energy band (high, low),
- rhythm descriptors, e.g.:
  - beat count (number of detected beats),
  - beat loudness (spectral energy computed on beat segments),
  - BPM value,
  - danceability,
  - onset rate (number of detected onsets per second),
- tonal descriptors, e.g.:
  - chord change rate,
  - key strength using diatonic profile.

Moreover, statistical descriptors can be used such as the mean, median, and standard deviation (stdev).

In our case, the descriptor extractor was based on the Essentia library. All descriptors listed above were used. For some descriptors, we used statistical data. Therefore, we had the following: mean values for 13 first MFCCs, more statistical values (such as the maximum, minimum, mean, median, stdev, and variance) for dynamic complexity, pitch salience, spectral complexity, spectral energy band (high and low), and beat loudness. In the case of dissonance, we used only the maximum, minimum, and stdev. We used a total of 54 parameters.

After selecting the appropriate model of the neural network, the process of evaluating the new generated recordings follows. It should be mentioned that our Critic is designed to evaluate the generated recordings as correct or incorrect, and this assignment to the appropriate category must be unambiguous and consistent. The main stages employed by our system are described in Figure 4.

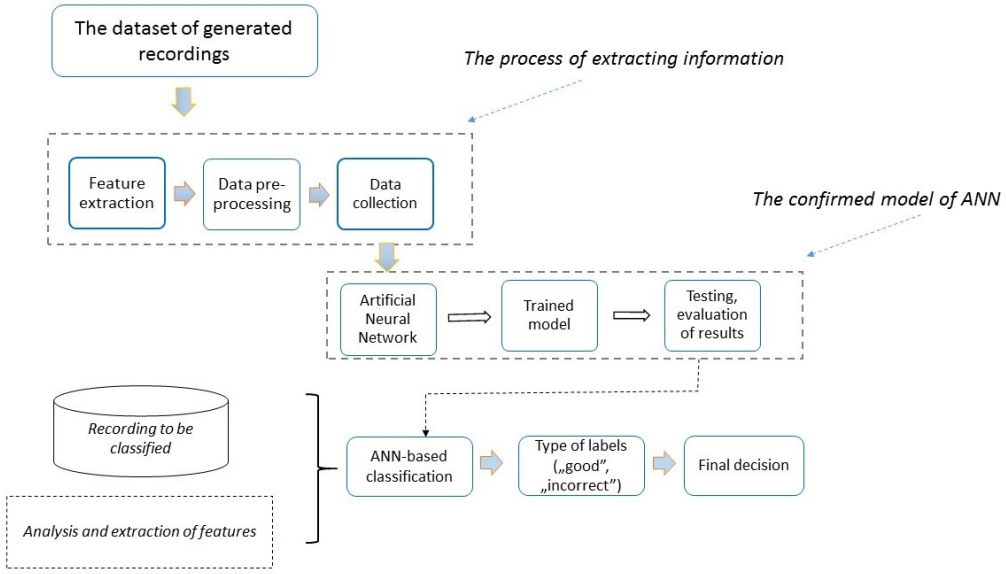

Figure 4. Main idea of the Critic-side.

Due to the computational complexity, classical neural networks (MLP, multilayer perceptron) were proposed for the Critic unit. They were taught on the basis of parametric information. Note that the use of mel-spectrograms or a combination of mel-spectrograms and parametric information would require the use of deep learning networks.

It was assumed that the input layer of the MLP consists of 54 neurons (taking into account various parameters and their selected statistical descriptors), and the output layer consists of 2 neurons (0—incorrect, 1—good). The neural networks are considered as MLP: 54-x-2, in which x denotes the number of neurons of the hidden layer. Depending on particular genres, the number of neurons in the hidden layer was different. In the case of relaxation music, the best network was the one containing 42 neurons in the hidden layer. For dance and electronic music, this number was equal to 35 and 22, respectively. The hyperbolic tangent activation function was used for neurons in the hidden layer, and the logistic function was used in the output layer. Moreover, the sum of squares is used as a function of the error, and nets are taught by the BFGS (Broyden-Fletcher-Goldfarb-Shanno) method.

3. Results and Discussion

3.1. Results of the Generator–Critic System Operation

The system based on the above discussed data representations and methods automatically produces varied musical works. Graphical representations of two examples, belonging to relaxation and electronic music, are presented in Figures 5–8.

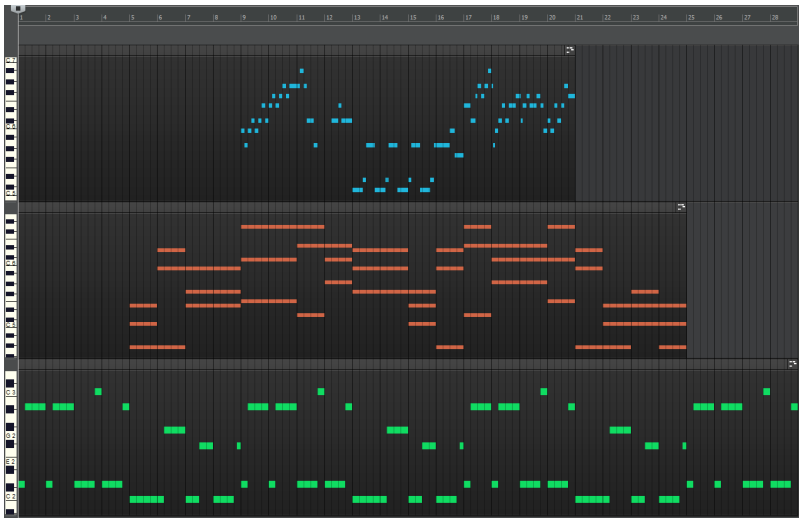

Figure 5. A piano roll view of generated relaxation music.

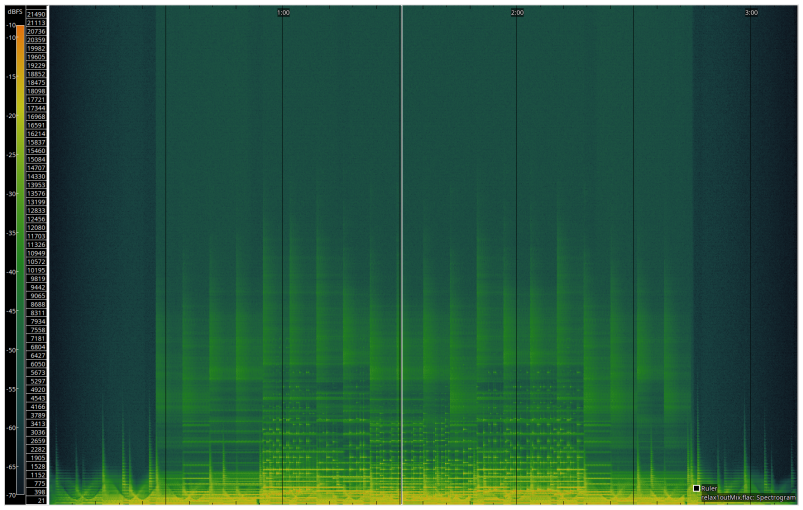

Figure 6. A spectrogram view of generated relaxation music.

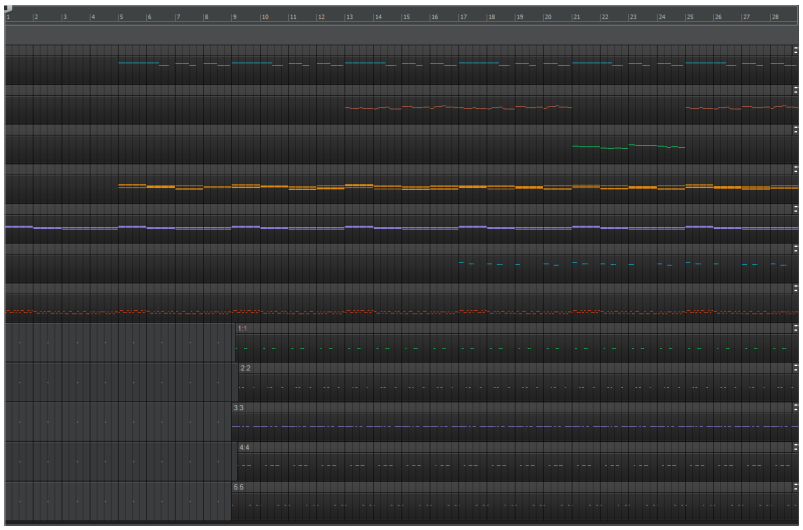

Figure 7. A piano roll view of generated electronic music (first 28 measures.)

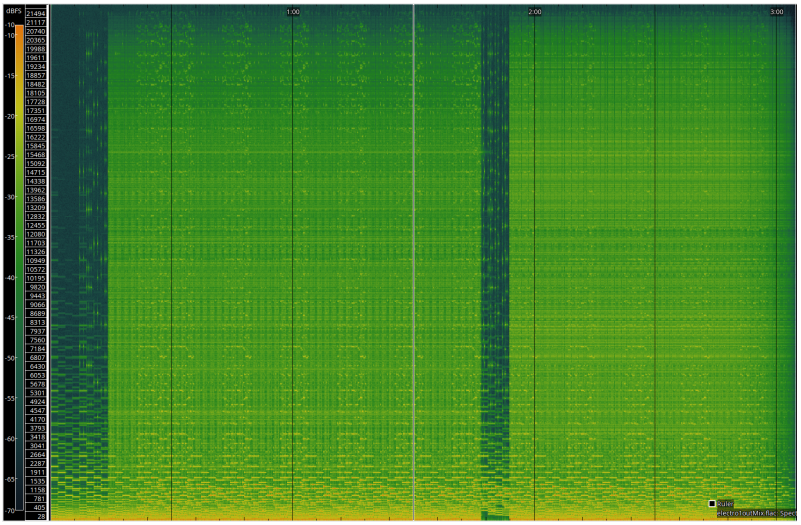

Figure 8. A spectrogram view of generated electronic music.

Relaxation music may consist of a small set of simple tracks (Figure 5). Much of its spectro-temporal complexity (Figure 6) is introduced at later stages through a selection of specific instruments and sounds and application of modulation effects.

Electronic music is more complex in the composition stage, with more tracks and more sophisticated forms (Figure 7). The modulation effects are less prominent compared to the variability already present in the actual score (Figure 8).

Another example may explain the lower level musical features, such as how motifs are processed into phrases and how phrases are used in the context of a whole song. In a simple relaxation song with a section length of two measures, the Generator produced an initial set of three motifs, as presented in Table 3. It is important to note that a song can use multiple musical scales, so a single interval value in a motif may represent different numbers of semitones depending on the actual scale and the scale step the interval is measured from (Figure 9).

Table 3. An exemplary set of motifs in a form used by the Generator to build phrases for the first lead track. A rhythm is expressed as a sequence of rhythmic values (2—half note, 4—quarter, 8—eight, and 16—sixteenth); the number of values equals the number of notes in a motif. Intervals represent pitch distances between subsequent notes, expressed as the number of steps on a selected musical scale (a negative value represents a movement downward).

Label	Rhythm	Intervals
mot1	8, 16, 16, 8	1, 2, 1
mot2	4, 8, 8, 4, 8	5, 1, 1, −2
mot3	8, 8, 2, 8	1, −1, −2

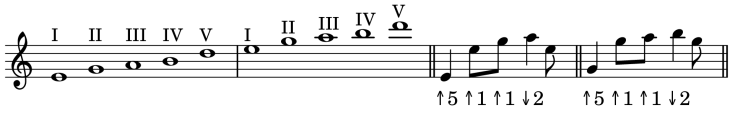

Figure 9. An example of the second motif from Table 3 (mot2) in a direct form (non-altered) in the "e" anhemitonic pentatonic scale (presented on the left through two octaves), starting from the first and the second step of the scale.

Using these three motifs, the Generator designed a number of phrases, three of which are presented in Figure 10, with motifs marked by dashed slurs. While designing phrases, the Generator rarely uses motifs directly but rather alters them, particularly with regards to intervals that can be inverted, diminished, or augmented, according to various rules to fit a

particular pitch range, scale, or other musical features. The rhythm can be altered as well, but in this case, it is visible only in skipping the ending of mot2 in (C). The GA takes care of including important melodic features, such as, e.g., a single climax (the highest pitch) in a phrase, or leaving the last note longer. The list of such features is much longer. Some other examples are avoiding excessive repetitions and preferring characteristic beginnings, such as double repetition or a large jump.

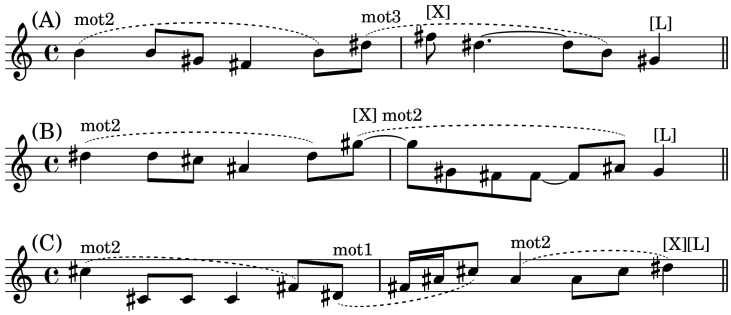

Figure 10. Three different phrases (**A**–**C**) produced by the Generator using motifs from Table 3. [X] marks a melody climax, [L] marks a longer ending note, and mot1-mot3 and dashed slurs mark the use of particular motifs.

Finally, a phrase is embedded in a section with the remaining tracks, as presented in Figure 11. Here, the accompanying layer is very basic, consisting of pads and bass tracks only, both using the simplest variant of long, standing chords. A prior section, without a lead melody, is also displayed because in the form selected for this particular song sections containing two lead voices (section *L1* containing the first voice, and section *L2* with the second voice) are interleaved with each other and with the section containing accompaniment only (section *Acc*), as follows: *L1-L2-Acc*. With each repetition of this sequence, phrases in sections *L1* and *L2* are altered.

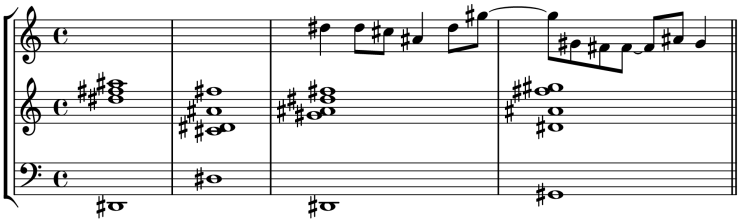

Figure 11. Phrase (B) from Figure 10 with accompanying tracks—chords and bass.

Our hybrid AI system may be seen as more versatile than systems based purely on deep learning techniques that need to be based upon particular examples and produce what can be considered variations of these examples. Our system allows the configuration of several key features to produce various genres and to adjust generated music according to user expectations.

Table 4 shows data obtained with the group of experts during the training phase of the Critic, i.e., assignments for generated recordings belonging to three genres: relaxation, dance, and electronic, each containing 2000 audio files.

Table 4. Number of correct and incorrect recordings for the three genres.

Genre	Correct (Class 1)	Incorrect (Class 2)
Relaxation	940	1060
Dance	659	1341
Electronic	311	1689

In order to determine the quality of the classification, the numbers of correctly classified positive and negative cases can be determined and marked, respectively, as TP (true positive) and TN (true negative) and incorrectly classified as FP (false positive) and FN (false negative). This is schematically presented in Table 5. We have seen significant differences in the classification results, relating to various genres of recordings. Out of 940 correct examples of relaxation music, 540 were correctly classified, accounting for about 57%. For dance music, 312 out of 659 correct examples were well classified, accounting for about 47%, and for electronic music, it was only 10% because 31 out of 311 correct recordings were appropriately classified. Referring to the incorrect class, we can see a better classification quality. For example, out of 1060 samples of relaxation music, 765 were properly classified, yielding about 72%. For dance and electronic music, it was about 89% and 99%, respectively. It needs to be pointed out that from a practical point of view, it is better to discard good music (classified as bad) and try to generate a new song again than to give a user a poorly generated piece as a correctly generated one.

Table 5. The number of correct and incorrect predictions for three types of genres.

Genre	Actual	Predicted Negative	Predicted Positive
Relaxation	Negative	TN = 765	FP = 400
	Positive	FN = 295	TP = 540
Dance	Negative	TN = 1199	FP = 347
	Positive	FN = 142	TP = 312
Electronic	Negative	TN = 1674	FP = 280
	Positive	FN = 15	TP = 31

The accuracy is one of the quality factors of a classification, and it refers to the percentage of correctly classified samples. It is equal to (TP + TN)/(TN + FP + FN + TP). Presented models of the classifiers achieved an accuracy equaling 0.65 for relaxation music, 0.75 for dance music, and 0.85 for electronic music. Another well-known quality factor is the recall (sensitivity), which is the proportion of actual positives that are correctly classified. This indicator is depicted as TP/(TP + FN), and it was similar for relaxation, dance, and electronic music. It was equal to 0.65, 0.69, and 0.67, respectively. Unfortunately, for these generated samples, we were not able to obtain better classification results. As one can see, the problem may exist due to a small number of collections of the assessed recordings. Moreover, in the considered data, there were not enough good recordings to build an accurate classifier. For example, for electronic music, only 311 out of 2000 were rated as good recordings by experts.

3.2. Auditory Evaluation

A listening test was carried out to evaluate the ability of the entire generator-critic system to produce music acceptable within boundaries of selected genres. A panel of listeners was given a task to evaluate if a given musical piece was a good example of a specific genre. A set of examples contained both fragments of musical recordings generated by our system and fragments of recordings produced by a human artist. Listeners were not informed about the origin of the evaluated recordings and were tasked with evaluating both the human- and computer-generated music under the same rules.

The system generated 50 recordings in each of the three genres: relaxation music, dance music, and electronic music. Out of these, five recordings were randomly selected from each genre. Accordingly, five recordings for each of the same genres were randomly selected from a repository of tagged music produced by human artists. Thus, a set of 30 recordings was selected, 10 in each of the three genres. Out of each of the 10 recordings within a genre, 5 were created by human artists and 5 were automatically produced by our system. In order to limit the duration of the test to avoid an effect of fatigue on the part of the listeners, only a 30 s fragment from the middle of each recording was extracted, with 2 s of fade-in and fade-out applied. Recordings were normalized to −3 dB FS. Within

each genre, a 10-element list of recordings was created, with 5 computer-generated and 5 human-created recordings randomly interleaved and arranged in random order.

The listeners used headphones in a silent room. They could listen to the recordings in any order, without a limit on the number of replays. Each recording was evaluated for fitness for a given genre using a six-grade scale, from bad (1) to poor (2), fair (3), good (4), very good (5), and excellent (6), as recommended in the EBU (European Broadcasting Union) document [3]. The listeners were also asked to state their age group, music production or music distribution experience, and any hearing problems.

Twenty listeners attended the test, 55% were between 20 and 29 years old, 36% were between 30 and 39 years old, and 9% were between 40 and 49 years old. In total, 55% had a professional experience with music production or distribution, 18% had an amateur experience, and 27% had minimal experience. No participants reported hearing loss.

The aim of the test was to determine if the system can produce music acceptable within specific genres; therefore, we decided to evaluate the quality of the recordings using the following filters for answers:

- accepted recordings, if human voted as fair, good, very good, or excellent,
- rejected recordings, if human voted as bad or poor.

As represented in Figure 12, the dance music generated by our system had the lowest level of acceptance (about 24%). It is interesting to observe that the recordings generated as relaxation music were largely accepted by the listeners (about 78%). Rejections of human-created recordings may be unexpected, but they can be explained by the procedure of selecting recordings for the test. In order to compare generated content with typical real-world counterparts, human-created recordings were randomly picked from a large commercial repository based on tags provided by authors or publishers to reflect the way they are tagged in real-use cases. Employing an external expert group to pick or tag musical excerpts might reduce rejections, but such a group—small by its nature—would be biased depending on their predominant experience with particular genres or stages of the music business. The obtained results reflect an actual problem with objective categorization of recordings into separate, well-defined genres, which affects the performance of the Critic part of our system, initially based on human ratings.

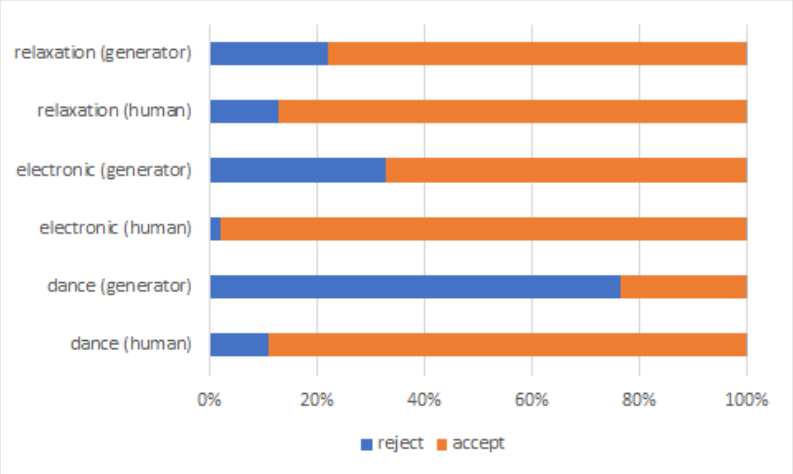

Figure 12. Evaluation of human-created and system-generated recordings by listener votes.

3.3. Legal Aspects

First, an explanation should be made that the recordings created by the system in question, from a legal point of view, consist of two components. The first one is a musical composition, which for the purpose of this paper can be described as notes. The second component is the musical recording, i.e., the sound layer. If the musical composition meets the requirements of the Copyright and Related Rights Act [43], it can be called a piece of

work. In turn, the first recording of a work or other acoustic phenomena is granted the status of a phonogram.

In the case of the phonogram, the issue is less problematic, as the Act [44] clearly indicates that the phonogram, which the producer holds the rights to, is the first recording of both a piece of work within the meaning of the Act and another acoustic phenomenon (it refers to all the recorded performance-related sounds, which may vary depending on a type of work). Therefore, the produced recording will receive copyright protection, which will be granted to its producer—the owner of the software.

The issue becomes more complicated when one attempts to determine the authorship of AI works. Currently, the doctrine generally accepts that AI by itself cannot be considered a creator within the meaning of the Copyright Law [45]. Similarly, as in the case of creations by animals (for example, the case of a picture taken by a monkey) [46], concepts suggesting to allow granting authorship to entities other than human beings were rejected.

While the generally held view is that only a human being can be regarded as a creator—and hence, only a product of human activity can be called a piece of work—several concepts [47] have been formulated with regard to granting the status of works for creations generated by artificial intelligence as well as assigning their authorship. In the opinion of the author of this paper, the most important concepts include the following: (1) granting authorship of a work to the creator of the software with the use of which that work was created, (2) granting authorship of a work to the "trainer" (a person who feeds the software data for machine learning), (3) granting authorship of a work to the owner of the software (the software such as the system for automatic music production presented in this manuscript) with the use of which that work was created, and (4) granting authorship to the software operator, who can determine the parameters of a future work.

The objective of this paper is not to solve the problem in question but to signal its existence. Regarding the concepts presented above, it seems that none of them would be fully adequate.

To conclude, in order to avoid doubts as to the copyright protection of works created by artificial intelligence, which may discourage prospective investors in such solutions, it is necessary for legislators to intervene and regulate these issues at the European and national level.

In the authors' opinion, it is worth considering a new concept, apart from those mentioned above, of a new category of related rights (to the works of artificial intelligence), which would be granted to the owner of an AI mechanism, and which in terms of protection and property rights would be equal to the right of the author of a work.

4. Conclusions

As shown in this paper, when planning to realize algorithmic music production with the use of AI methods, some aspects should be considered. Legal aspects related to creating recordings using AI algorithms and algorithmic composition used in our system are briefly described. AI algorithms are having an increasingly large impact on human creative and artistic endeavors. Legal aspects of creating musical compositions and recordings using AI algorithms are focused particularly on the issue of authorship and copyright protection of works created by artificial intelligence. Thus, the question of whether a composition created by artificial intelligence can be regarded as a piece of work, as understood in the Copyright Law, as well as the issue of its authorship, have not been resolved in Polish and UE legislation so far. Nevertheless, due to the increasing importance of artificial intelligence in the market, in recent years, this issue has gathered considerable interest in legal practice and from legal scholars. When considering the choice of a specific concept and its regulation, one cannot forget about its economic dimension. In the case of production of a musical recording by AI, the recording itself is protected due to the rights of the producer (usually the owner of the system) in the phonogram (subject to fixation of the sound); however, in the case of AI creating only musical compositions or an attempt to re-create

recordings of the created composition, the investor would face serious doubts regarding the protection of the created compositions.

The results come from our preliminary research on the simultaneous combination of generating recordings and their evaluation. Based on the results, one can see that the examined neural network is better at classifying incorrect recordings than good recordings. It can be seen that an AI-aided system that does not need to be "fed" external musical content, which would raise difficult legal questions, is able to generate complex and generally acceptable musical recordings. As confirmed by listening tests, in case of the relaxation music genre, the system based on presented methods produces largely acceptable results. On the other hand, dance music requires further refinements. It is worth mentioning that the time-consuming stage of evaluating more generated recordings by experts should be performed. Not all of the listeners in the evaluation panel were professionals, so recordings may not be tagged properly. We hope to continue the research to ensure the successful integration of the classifier and generator in automatic music production. The system components that we proposed constitute one of many possible ways to represent a working system. In future work, we want to investigate and examine other classifiers, including deep learning methods with an attention mechanism to obtain more proper music features. The richness of the music content needs techniques that take into account higher-level music features.

While legal issues may currently limit the use of systems, such as the one described here, to create and sell music, it still has many other perspective applications. It has three important advantages over the systems based on deep-learning techniques. One is the ability to strictly keep up to user-provided requirements, particularly the duration, but also genre and mood, always providing at least acceptable results. The other is avoiding legal problems with learning on other, possibly copyrighted music. Finally, it is very efficient, even in the current implementation (Python). Therefore, it can serve as a means to produce background music in real time for various spaces like lobbies, shops, etc. It can also become a part of a game-making or a movie-making tool. Once the legal problems mentioned earlier are solved at some point in time, other possible applications will emerge.

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

Funding: The article was funded by the research subsidy for AGH University of Krakow, subsidy numbers 16.16.120.773 and 16.16.130.942.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are unavailable due to privacy.

Conflicts of Interest: Andrzej Dąbrowski is Co-owner of Independent Digital sp. z o.o. and Bartłomiej Szadkowski is attorney of Independent Digital Sp. z o.o. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Aggarwal, C.C. *Data Classification: Algorithms and Applications*, 1st ed.; Chapman & Hall/CRC: Boca Raton, FL, USA, 2014.
2. Kotsiantis, S.; Zaharakis, I.; Pintelas, P. Machine learning: A review of classification and combining techniques. *Artif. Intell. Rev.* **2006**, *26*, 159–190. [CrossRef]

3. The European Broadcasting Union Document Tech 3286. Assessment Methods for the Subjective Evaluation of the Quality of Sound Programme Material—Music. 1997. Available online: https://tech.ebu.ch/publications/tech3286 (accessed on 3 October 2023).
4. Gungormusler, A.; Paterson-Paulberg, N.; Haahr, M. barelyMusician: An Adaptive Music Engine for Video Games. In Proceedings of the Audio Engineering Society Conference: 56th International Conference: Audio for Games, London, UK, 11–13 February 2015.
5. Williams, D.; Kirke, A.; Eaton, J.; Miranda, E.; Daly, I.; Hallowell, J.; Roesch, E.; Hwang, F.; Nasuto, S.J. Dynamic Game Soundtrack Generation in Response to a Continuously Varying Emotional Trajectory. In Proceedings of the Audio Engineering Society Conference: 56th International Conference: Audio for Games, London, UK, 11–13 February 2015.
6. Williams, D.; Hodge, V.; Gega, L.; Murphy, D.; Cowling, P.; Drachen, A. AI and Automatic Music Generation for Mindfulness. In Proceedings of the Audio Engineering Society Conference: 2019 AES International Conference on Immersive and Interactive Audio, York, UK, 27–29 March 2019.
7. Komosinski, M.; Szachewicz, P. Automatic species counterpoint composition by means of the dominance relation. *J. Math. Music* **2015**, *9*, 75–94. [CrossRef]
8. De Prisco, R.; Zaccagnino, G.; Zaccagnino, R. A Genetic Algorithm for Dodecaphonic Compositions. In Proceedings of the European Conference on the Applications of Evolutionary Computation, Aberystwyth, UK, 3–5 March 2011; pp. 244–253. [CrossRef]
9. Hiller, L.A., Jr.; Isaacson, L.M. Musical Composition with a High-Speed Digital Computer. *J. Audio Eng. Soc.* **1958**, *6*, 154–160.
10. Carnovalini, F.; Rodà, A. Computational Creativity and Music Generation Systems: An Introduction to the State of the Art. *Front. Artif. Intell.* **2020**, *3*, 14. [CrossRef]
11. Fernandez, J.; Vico, F. AI Methods in Algorithmic Composition: A Comprehensive Survey. *J. Artif. Intell. Res.* **2013**, *48*, 513–582. [CrossRef]
12. Donnelly, P.; Sheppard, J. Evolving Four-Part Harmony Using Genetic Algorithms. In Proceedings of the European Conference on the Applications of Evolutionary Computation, Aberystwyth, UK, 3–5 March 2011; pp. 273–282. _28 [CrossRef]
13. Mycka, J.; Żychowski, A.; Mańdziuk, J. Toward human-level tonal and modal melody harmonizations. *J. Comput. Sci.* **2023**, *67*, 101963. [CrossRef]
14. Briot, J.P.; Hadjeres, G.; Pachet, F.D. Deep Learning Techniques for Music Generation—A Survey. *arXiv* **2019**, arXiv:1709.01620.
15. Biswas, A.; Wennekes, E.; Wieczorkowska, A.; Laskar, R.H. (Eds.) *Advances in Speech and Music Technology. Computational Aspects and Applications*; Signals and Communication Technology; Springer: Cham, Switzerland, 2023. [CrossRef]
16. Ycart, A.; Benetos, E. Learning and Evaluation Methodologies for Polyphonic Music Sequence Prediction with LSTMs. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2020**, *28*, 1328–1341. [CrossRef]
17. Chen, J.; Pan, F.; Zhong, P.; He, T.; Qi, L.; Lu, J.; He, P.; Zheng, Y. An Automatic Method to Develop Music with Music Segment and Long Short Term Memory for Tinnitus Music Therapy. *IEEE Access* **2020**, *8*, 1. [CrossRef]
18. Huang, C.Z.A.; Vaswani, A.; Uszkoreit, J.; Shazeer, N.; Simon, I.; Hawthorne, C.; Dai, A.M.; Hoffman, M.D.; Dinculescu, M.; Eck, D. Music Transformer. *arXiv* **2018**, arXiv:1809.04281.
19. Min, J.; Liu, Z.; Wang, L.; Li, D.; Zhang, M.; Huang, Y. Music Generation System for Adversarial Training Based on Deep Learning. *Processes* **2022**, *10*, 2515. [CrossRef]
20. Neves, P.; Fornari, J.; Florindo, J. Generating music with sentiment using Transformer-GANs. *arXiv* **2022**, arXiv:2212.11134.
21. Jin, C.; Wang, T.; Liu, S.; Tie, Y.; Li, J.; Li, X.; Lui, S. A transformer-based model for multi-track music generation. *Int. J. Multimed. Data Eng. Manag.* **2020**, *11*, 36–54. [CrossRef]
22. Civit, M.; Civit-Masot, J.; Cuadrado, F.; Escalona, M. A systematic review of artificial intelligence-based music generation: Scope, applications, and future trends. *Expert Syst. Appl.* **2022**, *209*, 118190. [CrossRef]
23. Tzanetakis, G.; Cook, P. Musical Genre Classification of Audio Signals. *IEEE Trans. Speech Audio Process.* **2002**, *10*, 293–302. [CrossRef]
24. Lidy, T.; Rauber, A.; Pertusa, A.; Quereda, J.M.I. Improving Genre Classification by Combination of Audio and Symbolic Descriptors Using a Transcription Systems. In Proceedings of the ISMIR, Vienna, Austria, 23–27 September 2007; pp. 61–66.
25. Gan, J. Music Feature Classification Based on Recurrent Neural Networks with Channel Attention Mechanism. *Mob. Inf. Syst.* **2021**, *2021*, 1–10. [CrossRef]
26. Zhang, K. Music Style Classification Algorithm Based on Music Feature Extraction and Deep Neural Network. *Wirel. Commun. Mob. Comput.* **2021**, *2021*, 1–7. [CrossRef]
27. Ashraf, M.; Abid, F.; Din, I.U.; Rasheed, J.; Yesiltepe, M.; Yeo, S.F.; Ersoy, M.T. A Hybrid CNN and RNN Variant Model for Music Classification. *Appl. Sci.* **2023**, *13*, 1476. [CrossRef]
28. Nasrullah, Z.; Zhao, Y. Music Artist Classification with Convolutional Recurrent Neural Networks. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019; pp. 1–8. [CrossRef]
29. Laurier, C.; Grivolla, J.; Herrera, P. Multimodal Music Mood Classification Using Audio and Lyrics. In Proceedings of the 2008 Seventh International Conference on Machine Learning and Applications, San Diego, CA, USA, 11–13 December 2008; pp. 688–693. [CrossRef]
30. Seo, Y.S.; Huh, J.H. Automatic Emotion-Based Music Classification for Supporting Intelligent IoT Applications. *Electronics* **2019**, *8*, 164. [CrossRef]

31. Ferreira, P.; Limongi, R.; Favero, L.P. Generating Music with Data: Application of Deep Learning Models for Symbolic Music Composition. *Appl. Sci.* **2023**, *13*, 4543. [CrossRef]
32. Guo, Y.; Liu, Y.; Zhou, T.; Xu, L.; Zhang, Q. An automatic music generation and evaluation method based on transfer learning. *PLoS ONE* **2023**, *18*, e0283103. [CrossRef]
33. Gallagher, M. *The Music Tech Dictionary: A Glossary of Audio-Related Terms and Technologies*; Course Technology; Muska/Lipman: St. Clairsville, OH, USA, 2009.
34. Mido Webpage. Available online: https://mido.readthedocs.io/en/stable/ (accessed on 21 March 2024).
35. Pyo Webpage. Available online: https://pypi.org/project/pyo/ (accessed on 21 March 2024).
36. LilyPond Webpage. Available online: https://lilypond.org/ (accessed on 21 March 2024).
37. FluidSynth Webpage. Available online: https://www.fluidsynth.org/ (accessed on 21 March 2024).
38. SoundFont Technical Specification. Available online: http://www.synthfont.com/sfspec24.pdf (accessed on 21 March 2024).
39. SoX Webpage. Available online: https://sourceforge.net/projects/sox/ (accessed on 21 March 2024).
40. Engelbrecht, A.P. *Computational Intelligence: An Introduction*; John Wiley & Sons: Hoboken, NJ, USA, 2007. [CrossRef]
41. Schedl, M.; Hauger, D.; Urbano, J. Harvesting microblogs for contextual music similarity estimation: A co-occurrence-based framework. *Multimed. Syst.* **2014**, *20*, 693–705. [CrossRef]
42. Bogdanov, D.; Haro, M.; Fuhrmann, F.; Gómez, E.; Herrera, P. Content-based music recommendation based on user preference examples. In Proceedings of the ACM Conference on Recommender Systems. Workshop on Music Recommendation and Discovery (Womrad 2010), Barcelona, Spain, 26 September 2010 ; Volume 633.
43. Act of 4 February 1994 on Copyright and Related Rights (in Polish). Available online: http://www.prawoautorskie.gov.pl/media/download_gallery/D19940083Lj_19.07.pdf (accessed on 3 October 2023).
44. Article 94, Section 1 of the Act of 4 February 1994 on Copyright and Related Rights (in Polish). Available online: http://www.prawoautorskie.gov.pl/media/download_gallery/D19940083Lj_19.07.pdf (accessed on 3 October 2023).
45. Wojtczak, S.; Księżak, P. Copyright Law towards Artificial Intelligence (An Attempt at An Alternative View). *State Law (PańStwo Prawo)* **2021**, *2*, 21. (In Polish)
46. Guadamuz, A. The monkey selfie: Copyright lessons for originality in photographs and internet jurisdiction. *Internet Policy Rev.* **2016**, *5*, 1. [CrossRef]
47. Szpyt, K. The Use of Artificial Intelligence in Post-mortem Creativity and the Copyright of the Deceased Creator (in Polish). In *The Law of Artificial Intelligence (Polish: Prawo Sztucznej Inteligencji)* ; Lai, L., Świerczyński, M., Eds.; C.H. Beck: Warsaw, Poland, 2020; pp. 160–161.