

## О МОДУЛЬНОМ ПОДХОДЕ К МУЗЫКАЛЬНОМУ ТВОРЧЕСТВУ С ПОДДЕРЖКОЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Рогозинский Г. Г.<sup>1,2,4</sup>, доктор техн. наук, ✉ [gleb.rogozinsky@gmail.com](mailto:gleb.rogozinsky@gmail.com),  
[orcid.org/0000-0001-5698-2347](https://orcid.org/0000-0001-5698-2347)

Щекочихин А. В.<sup>4</sup>, разработчик, [ashekochikhin@gmail.com](mailto:ashekochikhin@gmail.com)

Малахов В. Е.<sup>2,3</sup>, аспирант, [vladyslavmalakho95@gmail.com](mailto:vladyslavmalakho95@gmail.com)

Горбунова И. Б.<sup>1</sup>, доктор пед. наук, профессор, [gorbunovaib@rgpu.spb.ru](mailto:gorbunovaib@rgpu.spb.ru),  
[orcid.org/0000-0003-4389-6719](https://orcid.org/0000-0003-4389-6719)

<sup>1</sup>Российский государственный педагогический университет им. А. И. Герцена,  
набережная реки Мойки, д. 48, 191186, Санкт-Петербург, Россия

<sup>2</sup>Институт проблем транспорта им. Н. С. Соломенко Российской Академии Наук,  
12ая линия В. О., д. 13, 199178, Санкт-Петербург, Россия

<sup>3</sup>Санкт-Петербургский государственный университет телекоммуникаций им. проф. М. А. Бонч-Бруевича,  
пр. Большевиков, 22 к. 1, 193232, Санкт-Петербург, Россия

<sup>4</sup>Омни Аудио, ул. Мери-Растилан, д. 23, пом. 17, 00980, Хельсинки, Финляндия

### Аннотация

В последнее время системы музыкального творчества с использованием искусственного интеллекта стремительно развиваются. Быстрый рост методов машинного обучения уже привел к появлению множества мощных решений в области синтеза звука, генеративных технологий, извлечения информации из музыки и других направлений. Доступность вычислительных ресурсов способствует прогрессу в разработке решений для обработки необработанного аудио. Декомпозиция задач может снижать сложность предлагаемых решений, обеспечивая повышение качества на каждом этапе производства.

**Ключевые слова:** музыкальные компьютерные технологии, модульность, музыкальное программирование, искусственный интеллект, моделирование.

**Цитирование:** Рогозинский Г. Г., Щекочихин А. В., Малахов В. Е., Горбунова И. Б. О модульном подходе к музыкальному творчеству с поддержкой искусственного интеллекта // Компьютерные инструменты в образовании. 2024. № 4. С. 59–69. doi:10.32603/2071-2340-2024-4-59-69

### 1. ВВЕДЕНИЕ

Методы, поддерживаемые искусственным интеллектом (ИИ), быстро развиваются в таких областях, как генерация музыки, импровизация, авторское право, оркестровка, поиск музыки и рекомендации плейлистов, ремикширование и продюсирование, музыкальный анализ, диджеинг, репетиции, обучение и выступления вживую. Анализируя развитие технологий компьютерной музыки с середины XX века, можно отметить, что предложенные решения зачастую становятся трендами благодаря соответствующим

основным технологиям, а не метрикам, специфичным для данной области. В настоящее время методы, основанные на машинном обучении (МО), находятся в центре внимания. Особенное внимание в этом плане уделяется сфере музыкального творчества и образования, в рамках которых наблюдается существенный спрос на использование новых методов и подходов к реализации музыкально-образовательного процесса на различных его уровнях (от общего до профессионального) и этапах (от начального до дополнительного профессионального и инклюзивного) [1–3]. Несмотря на значительный успех подходов, основанных на МО, в указанных областях, существуют следующие известные проблемы:

- крупные наборы данных и облачные вычислительные фермы недоступны для небольших исследовательских групп и конечных пользователей,
- неконтролируемое поведение предлагаемых инструментов («черный ящик»),
- высокие экономические и экологические затраты,
- культурная предвзятость в подходах, основанных на данных,
- сложность поддержки.

Мы предполагаем, что модульная архитектура может преодолеть перечисленные выше трудности.

Фундаментальные исследования, проводимые в области применения медиаинформационных систем и музыкально-компьютерных технологий, позволяют выделить цифровые технологии в музыке и саунд-дизайне как особую область знаний и базисную часть практически любой сферы деятельности в данном направлении, которое развивается на стыке музыкальной акустики, цифровой обработки сигналов, программной инженерии, разработка программного обеспечения на основе открытого исходного кода, моделирования процесса музыкального творчества и активно используется в системе современного музыкально-информационного и инженерно-технологического образования. Разработка специализированных музыкально-технологических дисциплин, таких как: «Технологии художественной обработки звука», «Музыкально-компьютерные технологии», «Музыкальные синтезаторы», «Музыкальная информатика», «Компьютерное музыкальное творчество», «Звукотембральное программирование», «Математические методы исследования в музыкологии», «Интеллектуальные системы каталогизации и анализа музыки народов мира», «Теория множеств и описание музыкальных объектов», «Теория информации и системный анализ музыки», «Саунд-дизайн и музыкальная звукорежиссура», «Компьютерная студия звукозаписи и основы работы в ней», «Компьютерное моделирование процесса музыкального творчества», требующих подготовки соответствующих учебных материалов (см., например, [4–6]), свидетельствует о становлении нового трансдисциплинарного научного и образовательного направления, обусловленного развитием современных музыкально-компьютерных технологий и формированием подходов к музыкальному творчеству с поддержкой элементов искусственного интеллекта.

Целью исследования является представление материалов по разработке элементов новой образовательной концепции, связанной с освоением блока профессиональных дисциплин в процессе подготовки специалистов в области цифровых технологий, обучающихся по образовательному направлению «Информационные системы и технологии».

## 2. ИНТЕЛЛЕКТУАЛЬНЫЕ МУЗЫКАЛЬНО-КОМПЬЮТЕРНЫЕ СИСТЕМЫ В ОБРАЗОВАНИИ

В настоящий момент на основании проводимых научно-педагогических исследований (см., например, 7–9) сотрудниками научно-методической лаборатории «Музыкально-

компьютерные технологии» Российского государственного педагогического университета им. А. И. Герцена разработаны и реализуются на базе лаборатории основные профессиональные образовательные программы подготовки бакалавров и магистров 09.03.02 «Информационные технологии в музыке и саунд-дизайне» и 09.04.02 «Цифровые технологии в музыке и саунд-дизайне» соответственно по направлению «Информационные системы и технологии», а также пять программ профессиональной переподготовки специалистов с высшим и средним профессиональным образованием, среди которых отметим «Информационные технологии в музыке и музыкальном образовании» и «Технологии создания и художественной обработки звуковой информации».

### 3. МОДУЛЬНЫЕ ПРИЛОЖЕНИЯ В ИИ-МУЗЫКЕ

Некоторые подходы с использованием модульных архитектур для творчества с поддержкой компьютера уже существовали ранее. Например, в [10] представлена общая модульная архитектура для компьютерных систем, предназначенных для выразительного музыкального исполнения.

Требования к модульному проектированию компьютерно-музыкальных систем могут быть следующими:

- модули должны быть специфичны для предметной области, а не для технологии, что позволяет менять базовое решение;
- модули должны иметь общий интерфейс для связи и конфигурации.

Рассмотрим некоторые примеры модулей и предложенные технологии для их реализации. Большая группа модулей предназначена для выполнения задач анализа и извлечения информации. Отслеживание гармонии и тактов может быть реализовано на основе [11] и [12] соответственно.

С помощью библиотеки *Omnizart* существует возможность выполнить задачи музыкальной транскрипции, то есть выделить и записать в символьном виде музыкальные инструменты, вокальную мелодию и гармонию из аудиофайла [13–15].

Другая группа модулей предназначена для решения задач распознавания формы: обнаружение припева и куплета, выделение аудиопревью и т. д. В [16] предложен алгоритм для выявления наиболее эмоциональной части трека, что критически важно для генерации аудиопревью. Модуль обнаружения припева может быть реализован на основе [17]. В [18] приводится сравнение алгоритмов обнаружения припева.

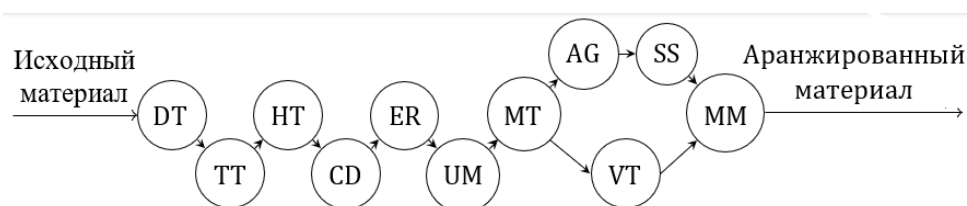
Очевидно, что музыка — это не просто сигнал в определенном диапазоне частот. Она, как правило, призвана передать слушателю определенное сообщение, следовательно, необходим модуль перцепционного и эмоционального анализа [19].

Модули сведения и мастеринга являются заключительными этапами в цепи обработки для большинства производственных сценариев. Представляется целесообразным разбить их на более мелкие части. Некоторые подходы, поддерживаемые ИИ для сведения и мастеринга музыки, можно найти в [20–23].

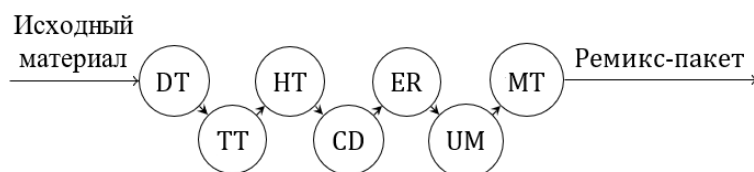
Рассмотрим решение сложных реальных задач в области аудиопроизводства с использованием комбинации предлагаемых модулей. На рис. 1 представлен пример трех таких модульных систем.

Первая система — Система аранжировки в определенном стиле. Она производит ре-аранжированную версию на основе исходной аудиозаписи. Цепь начинается с модулей анализа: отслеживания сильной доли (*Downbeat Tracking*, DT), отслеживания

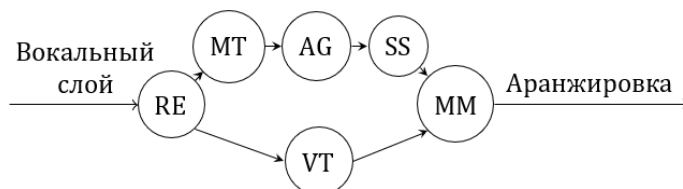
темпа (*Tempo Tracking*, TT) и отслеживания гармонии (*Harmony Tracking*, HT). Далее идет модуль обнаружения припева (*Chorus Detection*, CD), который выделяет наиболее значимую часть песни для дальнейшего аранжирования. Модуль ранжирования по эмоциональной окраске (*Emotional Ranking*, ER) используется для получения результатов анализа с точки зрения человеческого восприятия. Выбранный фрагмент должен быть разделен на составляющие для извлечения оригинальных вокальных и инструментальных партий с помощью модуля демикширования (*Un-Mix Module*, UM), а затем преобразован в символьную форму (например, в MIDI или в другую нотационную систему) с использованием модуля музыкальной транскрипции (*Music Transcription*, MT). После этого полученные данные управляют процессом генерации новой аранжировки в модуле *Arrangement Generation*, AG. Полученная аранжировка обрабатывается в модуле синтеза звука (*Sound Synthesis*, SS), затем, совмещенная с трансформированным вокалом, поступает на модуль сведения и мастеринга (*Mixing and Mastering*, MM).



(а) Система аранжировки в определенном стиле



(б) Система разделения на слои



(в) Система аранжировки

Рис. 1. Примеры архитектуры модульных систем компьютерной музыки

Вторая система — Система разделения на слои. Она создает пакет для будущих ремиксов, включающий аудиодорожки и MIDI, полученные на основе исходной композиции. Обработка начинается с тех же модулей анализа: DT, TT и HT. Далее идут модули обнаружения припева и ранжирования по эмоциональной окраске для определения самых значимых частей и запоминающихся музыкальных фрагментов. Затем выбранные части обрабатываются модулем демикширования (UM) для извлечения изолированных вокальных партий и паттернов ударных. Также используется модуль транскрипции для создания всех необходимых MIDI-дорожек.

Наконец, рассмотрим простую систему аранжировки, основанную на записанной пользователем вокальной партии. Пользователь записывает вокальную партию в модуле записи (*Record Module*, RE) под метроном для обеспечения синхронизации с заданным темпом. Одна часть системы предназначена для перевода записи голоса в MIDI в модуле MT, затем для создания символьной версии аранжировки в модуле AG на основе записанной мелодии. Далее идет модуль синтеза звука (SS). В параллельной части вокал тщательно подстраивается, выравнивается и обрабатывается в модуле трансформации вокала (*Vocal Transformation*, VT). В итоге слои композиции поступают в модули сведения и мастеринга.

Предложенные системы могут легко расширяться, модернизироваться и обеспечивать необходимый пользовательский контроль на каждом этапе.

#### 4. МОДЕЛЬ КОММУТАЦИИ МОДУЛЕЙ

Пусть существуют два модуля, первый из которых имеет входной терминал, а другой — выходной. Входным терминалам поставим в соответствие элемент  $\odot$ , а выходным —  $\ominus$  из множества  $\varepsilon = \{\odot, \ominus\}$  с заданной на нем унарной операцией инверсии  $\overline{\odot} \rightarrow \ominus$ ,  $\overline{\ominus} \rightarrow \odot$ .

Каждому типу терминала поставим в соответствие один и только один элемент из множества типов музыкальной информации  $c = \{c^A, c^M, c^S\}$ , где  $c^A$  — звуковые сигналы,  $c^M$  — модуляционные сигналы,  $c^S$  — символьная информация. Отсутствие музыкальной информации в терминале обозначим через  $c = \{\emptyset\}$ .

Еще одно понятие, которое нам необходимо ввести — интерфейс  $\pi$ , определяющий способ соединения отдельных модулей, а также различные подсистемы между собой. Интерфейс мы должны понимать здесь в широком смысле, не ограничиваясь только физическими интерфейсами, уровнями сигналов, или протоколами передачи.

В таком случае каждый терминал  $z \in Z$  может быть формализован в виде тройки  $\langle \varepsilon, c, \pi \rangle$ , где  $\varepsilon \in E, c \in C, \pi \in \Pi$ .

Тогда условием соединения двух терминалов  $z_1, z_2 \in Z$  будет являться

$$\varepsilon_1 \neq \varepsilon_2, \quad c_1 = c_2, \quad \pi_1 = \pi_2.$$

Под функцией соединения  $Q$  двух терминалов  $z_1, z_2 \in Z$  по интерфейсу  $\pi \in \Pi$  примем

$$Q : Z \times Z \times \Pi \rightarrow \{0, 1\},$$

$Q = 1$  означает, что соединение присутствует,  $Q = 0$  означает обратную ситуацию.

В качестве иллюстрации на рисунке 2 представлена система из трех модулей 1–3, модуль 1 имеет входную связь с внешней средой, модуль 3 имеет выходную связь с внешней средой. Терминалы  $z$  каждого модуля имеют два индекса. Верхний индекс определяет принадлежность терминала к одноименному модулю, нижний является порядковым номером данного терминала.

Модуль 1 совершает согласование немзыкальных входных данных  $X$  с системой, в результате чего на его выходном терминале  $z_2^1$  появляется музыкальная информация в виде звукового сигнала. С помощью интерфейса типа  $\pi_1$  терминал  $z_2^1$  соединен с модулем 2 (входной терминал  $z_1^2$ ). По этому соединению передаются сигналы типа  $c^A$  (звуковые). Модуль 2 имеет связь с модулем 3 посредством интерфейса типа  $\pi_2$ , и эта связь используется для передачи символьных сообщений  $c^S$ . Модуль 3 с помощью интерфейса типа  $\pi_3$  передает во внешнюю среду звуковые сигналы через выходной терминал  $z_2^3$ .

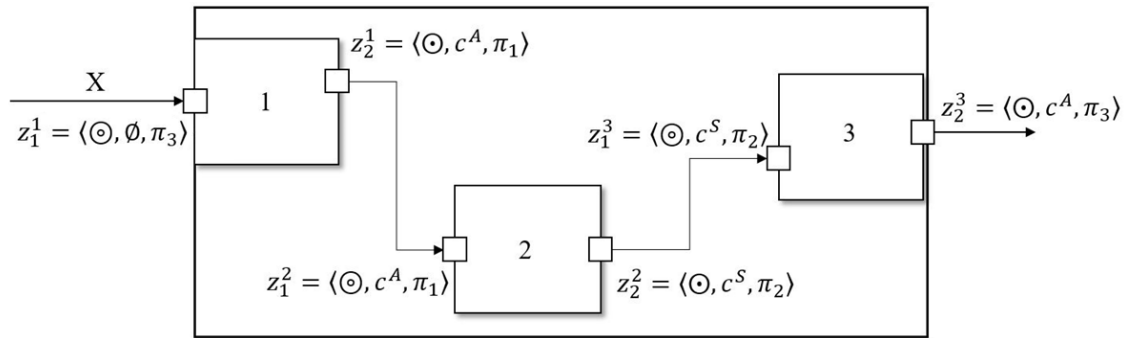


Рис. 2. Обозначения в модульной системе

## 5. CSOUND КАК ЯДРО ЗВУКОВОГО СИНТЕЗА

Язык музыкального программирования Csound [24] естественным образом рассматривается как основа универсального движка систем звукового синтеза и обработки сигналов. Csound обладает открытым API, что позволяет управлять как всем звуковым движком, так и отдельными инструментами. Он может запускаться внутри Python, одного из самых распространенных на сегодняшний день языков, используемых в технологиях ИИ.

Пример простого секвенсора (instr Pulse), управляющего субтрактивным синтезатором (instr 2) со случайным выбором частоты среза резонансного ФНЧ (moogvcf2) из диапазона 300–1300 Гц, приведен на рисунке 3.

```

sr = 44100                                ; установка частоты дискретизации
0dbfs = 1.0                               ; нормировочное значение амплитуды
nchnls = 2                                ; установка кол-ва выходных каналов

alwayson "Pulse"                           ; непрерывный режим инструмента
instr Pulse                                ; название инструмента
kclk metro 1                               ; метроном с частотой 1 Гц
if kclk != 1 goto Halt                    ; условный переход при отсутствии клика
event "i", 2, 0, 1                        ; запуск очередного звука по выполнению условия
Halt:                                     ; метка перехода
endin                                     ; завершение описания инструмента

instr 2                                    ; название инструмента
icut random 300, 1300                     ; случайное значение от 300 до 1300
kEnv linseg 0, 0.1, 1, 0.5, 0             ; амплитудная огибающая из 2х стадий
a1 rand 0.5 * kEnv                         ; генерация белого шума
a2 moogvcf2 a1, 100 + kEnv * icut, 0.8    ; фильтрация
outs a2, a2                               ; вывод звука
endin                                     ; завершение описания инструмента

```

Рис. 3. Пример кода на языке Csound

Современные версии Csound поддерживают синтез и обработку звука в реальном времени, поэтому различные модули на основе Csound подходят, например, для систем, используемых в живых выступлениях, для систем генерации алгоритмических звуковых ландшафтов и других подобных решений. С развитием технологий связи, таких как WebRTC, модульная система может стать распределенной.

Кроме того, Csound хорошо документирован и продолжает развиваться благодаря усилиям международного сообщества. Открытость исходного кода Csound, непрерывное совершенствование, поддержка различных платформ, включая мобильные, обширный ин-



струментал, поддержка протоколов MIDI и OSC, обеспечивающих связность с другими программными и аппаратными системами компьютерной музыки, наличие справочной литературы на нескольких языках — все это делает язык музыкального программирования *Csound* идеальным решением для изучения музыкального программирования в рамках профильных специальностей высшей школы.

## 6. МОДУЛЬНЫЙ ИНТЕЛЛЕКТУАЛЬНЫЙ АРАНЖИРОВЩИК В СТИЛЕ ТЕХНО

Приведем пример системы-аранжировщика в стиле *техно*, основанного на несколько упрощенной архитектуре системы, описанной ранее. Он был разработан с использованием *Csound* в качестве ядра звукового синтеза. Результаты переработки предварительного просмотра песни в стиле *техно* в формате MIDI и окончательные звуковые результаты находятся по ссылке [https://github.com/ashekochikhin/modular\\_pub](https://github.com/ashekochikhin/modular_pub).

Модули анализа были разработаны с использованием библиотеки *librosa* для отслеживания темпа, *Omnizart* для отслеживания гармонии и *madmom* [25] для отслеживания сильных долей. Обнаружение припева было реализовано с помощью алгоритма *PyChorus* [26]. Библиотека *Open-Unmix* использовалась для выделения вокала из оригинального припева песни. Модуль *Pyrubberband* использовался для растягивания (стретчинга) вокала по времени. Модуль музыкальной транскрипции был реализован с использованием указанной версии алгоритма *Omnizart*. Модель генерации аранжировки в стиле *техно* использует систему правил (rule-based system). Модуль синтеза на основе *Csound* включает в себя сэмплы электронных ударных инструментов и FM-синтезаторы.

## 7. ЗАКЛЮЧЕНИЕ

Модульный подход к музыкальному творчеству с поддержкой ИИ является гибким, расширяемым и дружелюбным к новым технологиям. Независимая разработка различных частей систем, поддерживаемых ИИ, создает вызовы для реализации модулей. Мы рассматриваем следующие направления развития языков компьютерной музыки для дальнейшего совершенствования модульных систем:

- с развитием интерактивных компонентов мета- и дополненной реальности, генерация и синтез звуковых пространств потребуют больших вычислительных ресурсов, поэтому интеграция с DSP, CUDA или другими специализированными вычислительными системами кажется наиболее целесообразной;
- расширение API языков компьютерной музыки, особенно для веб-сред и сред обработки данных, представляется приоритетным.

Разработанные научные и образовательные направления являются действенной основой для формирования новой сферы музыкальной культуры, ярким образом иллюстрирующей непосредственную связь знаний в сфере науки о музыке и области компьютерных наук, которые используют в своей работе специалисты в сфере музыкальной информатики, компьютерного музыкального творчества, цифровых искусств, медиамузыки, компьютерной музыки, музыкального и звукотембрального программирования, медиаинформационных систем и технологий.

## Список литературы

1. Gorbunova I. The Integrative Model for the Semantic Space of Music and a Contemporary Musical Educational Process: The Scientific and Creative Heritage of Mikhail Borisovich Ignatyev // Educação. 2021. Vol. 46. P. 1–23. doi:10.5902/1984644453329

2. Горбунова И. Б., Заливадный М. С. Трансдисциплинарный подход к изучению музыкальных явлений: теория информации и её воздействие на различные области музыкознания // Проблемы музыкальной науки. 2024. № 2. С. 180–199. doi:10.56620/2782-3598.2024.2.180-199
3. Gorbunova I., Govorova A. Music Computer Technologies in Informatics and Music Studies at Schools for Children with Deep Visual Impairments: From the Experience // Lecture Notes in Computer Science. Proceedings. Springer. 2018. P. 381–389. doi:10.1007/978-3-030-02750-6\_29
4. Бажукова Е. Н., Горбунова И. Б., Заливадный М. С., Чибирёв С. В. Музыкальная информатика: учебное пособие. СПб.: Лань, Планета музыки, 2023. 208 с.
5. Горбунова И. Б. Информационные технологии в музыке. Кн. 1: Музыкальные синтезаторы: учебное пособие. М.: ЛЕНАНД, URSS, 2023. 200 с.
6. Горбунова И. Б. Информационные технологии в музыке. Кн. 2: Музыкальные синтезаторы: учебное пособие. М.: ЛЕНАНД, URSS, 2024.
7. Gorbunova I. B., Chibirev S. V. Modeling the Process of Musical Creativity in Musical Instrument Digital Interface Format // Opcion. 2019. Vol. 35. No. Special Issue 22. P. 392–409.
8. Chibirev S., Gorbunova I. Computer Modeling in Musical Creative Work: An Interdisciplinary Research Example // Lecture Notes in Networks and Systems. 2022. Vol. 345. P. 474–483. doi:10.1007/978-3-030-89708-6\_40
9. Горбунова И. Б. и др. Опыт компьютерного моделирования произведений музыкального творчества на основе комплексного исследования закономерностей музыки // Музыка, математика, информатика: комплексная модель семантического пространства музыки: монография. СПб.: Лань, Планета музыки. 2024. С. 159–252.
10. Kirke A., Miranda E. R. Performance Creativity in Computer Systems for Expressive Performance of Music // Handbook of Artificial Intelligence for Music. Springer, Cham. 2021.
11. Chang C.-C., Su L. Beast: Online Joint Beat and Downbeat Tracking Based on Streaming Transformer // Proc. of ICASSP 2024 — 2024 IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP). 2024. P. 396–400. doi:10.1109/ICASSP48485.2024.10446611
12. Boeck S., Krebs F., Widmer G. Joint Beat and Downbeat Tracking with Recurrent Neural Networks // Proceedings of the 17th International Society for Music Information Retrieval Conference (ISMIR). 2016. P. 255–261.
13. Hsu J.-Y., Su L. VOCANO: A Note Transcription Framework for Singing Voice in Polyphonic Music // Proc. International Society of Music Information Retrieval Conference (ISMIR). 2021. P. 295–300.
14. Wei I.-C., Wu C.-W., Su L. Improving Automatic Drum Transcription Using Large Scale Audio-to-Midi Aligned Data // Proc. of 2021 IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP). 2021. P. 246–250.
15. Wu Y.-T., Chen B., Su L. Multi-instrument Automatic Music Transcription with Self-attention-based Instance Segmentation // Proc. of IEEE/ACM Transactions on Audio, Speech, and Language. 2020. Vol. 28. P. 2796–2809.
16. Huang Y.-S., Chou S.-Y., Yang Y.-H. Pop Music Highlighter: Marking the Emotion Keypoints // Transactions of the International Society for Music Information Retrieval. 2018. Vol. 1, № 1. P. 68–78. doi:10.5334/tismir.14
17. Wang J.-C., Smith J. B.L., Chen J., Song X., Wang Y. Supervised Chorus Detection for Popular Music Using Convolutional Neural Network and Multi-Task Learning // Proc. of ICASSP 2021 — 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2021. P. 566–570. doi:10.1109/icassp39728.2021.9413773
18. Nieto O., Bello J. P. Systematic Exploration of Computational Music Structure Research // Proc. of the 17th International Society for Music Information Retrieval Conference (ISMIR). New York City, NY, USA. 2016.
19. Pyrovolakis K., Tzouveli P., Stamou G. Mood Detection Analyzing Lyrics and Audio Signal Based on Deep Learning Architectures // Proc. of 2020 25th International Conference on Pattern Recognition (ICPR). 2021. P. 9363–9370.
20. Aker O. AI-Assisted Music Mastering: An Exploration of Human and AI Practices in Contemporary Music Production // Understanding Generative AI in a Cultural Context. 2024. P. 17–50. doi:10.4018/979-8-3693-7235-7.ch002
21. Koszewski D. et al. Automatic music signal mixing system based on one-dimensional Wave-U-Net autoencoders // EURASIP Journal on Audio, Speech, and Music Processing. 2023. № 1. P. 1–17. doi:10.1186/s13636-022-00266-3
22. Othman A. ChatGPT + SUNO: A Beginner's Guide to AI-Assisted Songwriting and Music Production. 2024. doi:10.13140/RG.2.2.33670.48960.



23. Li Y., Liu X. An Intelligent Music Production Technology Based on Generation Confrontation Mechanism // Computational Intelligence and Neuroscience. 2022. P. 1–10. doi:10.1155/2022/5083146
24. Stojak S., Hofmann A. (eds.). Proceedings of the 7th International Csound Conference. Vienna: University of Music and Performing Arts Vienna, 2024. doi:10.21939/icsc2024.
25. Boeck S., Korzeniowski F., Schluete, J., Krebs F. Widmer G. madmom: A New Python Audio and Music Signal Processing Library // Extended abstracts for the Late-Breaking Demo Session of the 17th International Society for Music Information Retrieval Conference. 2016. P. 1174–1178. doi:10.1145/2964284.2973795
26. Jayaram V. Finding Choruses in Songs with Python [Электронный ресурс]. URL: <https://towardsdatascience.com/finding-choruses-in-songs-with-python-a925165f94a8> (дата обращения: 27.12.2024).

Поступила в редакцию 15.12.2024, окончательный вариант — 27.12.2024.

**Рогозинский Глеб Гендрихович**, доктор техн. наук, старший научный сотрудник, научно-методическая лаборатория «Музыкально-компьютерные технологии», РГПУ; Ведущий научный сотрудник лаборатории проблем экологии транспортных систем, ИПТ РАН, Санкт-Петербург, ✉ [gleb.rogozinsky@gmail.com](mailto:gleb.rogozinsky@gmail.com)

**Щекочихин Алексей Викторович**, разработчик, Омни Аудио, Финляндия, [ashekochikhin@gmail.com](mailto:ashekochikhin@gmail.com)

**Малахов Владислав Евгеньевич**, аспирант, кафедра Информатики и компьютерного дизайна СПбГУТ, Санкт-Петербург, [vladyslavmalakho95@gmail.com](mailto:vladyslavmalakho95@gmail.com)

**Горбунова Ирина Борисовна**, доктор педагогических наук, профессор, руководитель научно-методической лаборатории «Музыкально-компьютерные технологии», Санкт-Петербург, [gorbunovaib@rgpu.spb.ru](mailto:gorbunovaib@rgpu.spb.ru)

---

Computer tools in education, 2024

№ 4: 59–69

<http://cte.eltech.ru>

doi:10.32603/2071-2340-2024-4-59-69

## On Modularity in AI-Aided Music Systems

Rogozinski G. G.<sup>1,2,4</sup>, Doctor Sc., ✉ [gleb.rogozinsky@gmail.com](mailto:gleb.rogozinsky@gmail.com),  
[orcid.org/0000-0001-5698-2347](https://orcid.org/0000-0001-5698-2347)

Shchekochikhin A. V.<sup>4</sup>, Developer, [ashekochikhin@gmail.com](mailto:ashekochikhin@gmail.com)

Malakhov V. M.<sup>2,3</sup>, Postgraduate, [vladyslavmalakho95@gmail.com](mailto:vladyslavmalakho95@gmail.com)

Gorbunova I. B.<sup>1</sup>, Doctor Sc., Full Professor, [gorbunovaib@rgpu.spb.ru](mailto:gorbunovaib@rgpu.spb.ru),  
[orcid.org/0000-0003-4389-6719](https://orcid.org/0000-0003-4389-6719)

<sup>1</sup> Herzen State Pedagogical University of Russia,  
48 Moika river emb., 191186, Saint Petersburg, Russia

<sup>2</sup> Solomenko Institute of Transport Problem of Russian Academy of Sciences,  
13 12th lane V. O., 199178, Saint Petersburg, Russia

<sup>3</sup> The Bonch-Bruевич St. Petersburg State University of Telecommunications,  
22-1 Bolshevikov Pr., 193232, Saint Petersburg, Russia

<sup>4</sup> Omni Audio Oy, Meri-Rastilan tie 23 B 17, 00980, Helsinki, Finland

### Abstract

Recently, music creation systems utilizing artificial intelligence have been rapidly advancing. The swift growth of machine learning methods has already led to the emergence of numerous powerful solutions in the fields of sound synthesis, generative technologies, music information retrieval, and other areas. The availability of computational resources is driving progress in the development of solutions for processing raw audio. Decomposing tasks can reduce the complexity of proposed solutions, ensuring improved quality at each stage of production.

**Keywords:** *computer music technologies, modularity, music programming, artificial intelligence, modeling.*

**Citation:** G. G. Rogozinski, A. V. Shchekochikhin, V. M. Malakhov, and I. B. Gorbunova, "On Modularity in AI-Aided Music Systems," *Computer tools in education*, no. 4, pp. 59–69, 2024 (in Russian); doi:10.32603/2071-2340-2024-4-59-69

### References

1. I. Gorbunova, "The integrative model for the semantic space of music and a contemporary musical educational process: the scientific and creative heritage of Mikhail Borisovich Ignatyev," *Educação*, vol. 46, pp. 1–23, 2021; doi:10.5902/1984644453329
2. I. B. Gorbunova and M. S. Zalivadny, "The Trans-Disciplinary Approach to the Study of Musical Phenomena: The Information Theory and Its Impact on Various Fields of Musicology," *Problemy muzykal'noi nauki / Music Scholarship*, no. 2, pp. 180–199, 2024 (in Russian); doi:10.56620/2782-3598.2024.2.180-199
3. I. Gorbunova and A. Govorova, "Music Computer Technologies in Informatics and Music Studies at Schools for Children with Deep Visual Impairments: From the Experience," *Informatics in Schools. Fundamentals of Computer Science and Software Engineering*, pp. 381–389, 2018; doi:10.1007/978-3-030-02750-6\_29
4. E. N. Bazhukova et al., *Muzykal'naya informatika: uchebnoe posobie* [Musical informatics: a tutorial], St. Petersburg, Russia: Lan'. Planeta muzyki, 2023 (in Russian).
5. I. B. Gorbunova, *Informatsionnye tekhnologii v muzyke. Kn. 1: Muzykal'nye sintezatory: uchebnoe posobie* [Information Technology in Music. Book 1: Music Synthesizers: A Tutorial], Moscow: LENAND, URSS, 2024 (in Russian).
6. I. B. Gorbunova, *Informatsionnye tekhnologii v muzyke. Kn. 2: Muzykal'nye sintezatory: uchebnoe posobie* [Information Technology in Music. Book 1: Music Synthesizers: A Tutorial], Moscow: LENAND, URSS, 2024 (in Russian).
7. I. B. Gorbunova and S. V. Chibirev, "Modeling the Process of Musical Creativity in Musical Instrument Digital Interface Format," *Opcion*, vol. 35, no. Special Issue 22, pp. 392–409, 2019.
8. S. Chibirev and I. Gorbunova, "Computer Modeling in Musical Creative Work: An Interdisciplinary Research Example," *Lecture Notes in Networks and Systems*, vol. 345, pp. 474–483, 2022; doi:10.1007/978-3-030-89708-6\_40
9. I. B. Gorbunova et al., "Opyt komp'yuternogo modelirovaniya proizvedenii muzykal'nogo tvorchestva na osnove kompleksnogo issledovaniya zakonornostei muzyki" [Experience of computer modeling of musical works based on a comprehensive study of musical patterns], in *Muzyka, matematika, informatika: kompleksnaya model' semanticheskogo prostranstva muzyki: monograph*, I. B. Gorbunova et al. eds., pp. 159–252, St. Petersburg, Russia: Lan'. Planeta muzyki, 2024 (in Russian).
10. A. Kirke and E. R. Miranda, "Performance Creativity in Computer Systems for Expressive Performance of Music," E. R. Miranda ed., in *Handbook of Artificial Intelligence for Music*, Springer, Cham., 2021.
11. C.-C. Chang and L. Su, "Beast: Online Joint Beat and Downbeat Tracking Based on Streaming Transformer," in *Proc. of ICASSP 2024 — 2024 IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 396–400, 2024; doi:10.1109/icassp48485.2024.10446611
12. S. Boeck, F. Krebs, and G. Widmer "Joint Beat and Downbeat Tracking with Recurrent Neural Networks," in *Proc. of the 17th International Society for Music Information Retrieval Conference (ISMIR)*, pp. 255–261, 2016.
13. J.-Y. Hsu and L. Su, "VOCANO: A Note Transcription Framework for Singing Voice in Polyphonic Music," in *Proc. of the Int. Society of Music Information Retrieval Conference (ISMIR)*, pp. 295–300, 2021.
14. I.-C. Wei, C.-W. Wu, and L. Su, "Improving Automatic Drum Transcription Using Large Scale Audio-to-Midi Aligned Data," in *Proc. of 2021 IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 246–250, 2021.
15. Y.-T. Wu, B. Chen, and L. Su, "Multi-Instrument Automatic Music Transcription With Self-Attention-Based Instance Segmentation," in *Proc. of IEEE/ACM Transactions on Audio, Speech, and Language*, vol. 28, pp. 2796–2809, 2020, doi:10.1109/taslp.2020.3030482

16. Y.-S. Huang, S.-Y. Chou, and Y.-H. Yang, "Pop Music Highlighter: Marking the Emotion Keypoints," *Transactions of the International Society for Music Information Retrieval*, vol. 1, no. 1, pp. 68–78, 2018; doi:10.5334/tismir.14
17. J.-C. Wang et al., "Supervised Chorus Detection for Popular Music Using Convolutional Neural Network and Multi-Task Learning," in *Proc. of ICASSP 2021 — 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 566–570, 2021; doi:10.1109/icassp39728.2021.9413773
18. O. Nieto and J. P. Bello, "Systematic Exploration of Computational Music Structure Research," in *Proc. of the 17th International Society for Music Information Retrieval Conference (ISMIR)*, New York City, NY, USA, 2016.
19. K. Pyrovolakis, P. Tzouveli, and G. Stamou, "Mood Detection Analyzing Lyrics and Audio Signal Based on Deep Learning Architectures," in *Proc. of 2020 25th International Conference on Pattern Recognition (ICPR)*, pp. 9363–9370, 2021.
20. O. Aker, "AI-Assisted Music Mastering: An Exploration of Human and AI Practices in Contemporary Music Production," *Understanding Generative AI in a Cultural Context*, pp. 17–50, 2024; doi:10.4018/979-8-3693-7235-7.ch002
21. D. Koszewski et al., "Automatic music signal mixing system based on one-dimensional Wave-U-Net autoencoders," *EURASIP Journal on Audio, Speech, and Music Processing*, no. 1, pp. 1–17, 2023; doi:10.1186/s13636-022-00266-3
22. A. Othman, *ChatGPT + SUNO: A Beginner's Guide to AI-Assisted Songwriting and Music Production*, 2024; doi:10.13140/RG.2.2.33670.48960
23. Y. Li and X. Liu, "An Intelligent Music Production Technology Based on Generation Confrontation Mechanism," *Computational Intelligence and Neuroscience*, pp. 1–10, 2022; doi:10.1155/2022/5083146
24. S. Stojak and A. Hofmann, eds., *Proceedings of the 7th International Csound Conference*, Vienna: University of Music and Performing Arts Vienna, 2024; doi:10.21939/icsc2024
25. S. Boeck et al., "madmom: A New Python Audio and Music Signal Processing Library," in *Extended abstracts for the Late-Breaking Demo Session of the 17th International Society for Music Information Retrieval Conference*, pp. 1174–1178, 2016; doi:10.1145/2964284.2973795
26. V. Jayaram, "Finding Choruses in Songs with Python," in *towardsdatascience.com*, 2024. [Online]. Available: <https://towardsdatascience.com/finding-choruses-in-songs-with-python-a925165f94a8>

*Received 15-12-2024, the final version — 27-12-2024.*

**Gleb Rogozinski, Doctor Sc. (Tech.), Senior Researcher, Computer Music Labs, Herzen University; Lead Researcher, Laboratory of Ecological Problems of Transport Systems, IPT RAS, Saint Petersburg, ✉ [gleb.rogozinsky@gmail.com](mailto:gleb.rogozinsky@gmail.com)**

**Aleksei Shchekochikhin, Developer, Omni Audio Oy, Finland, [ashekochikhin@gmail.com](mailto:ashekochikhin@gmail.com)**

**Vladislav Malakhov, Postgraduate, Dept. Informatics and Computer Design, Saint Petersburg, [vladyslavmalakho95@gmail.com](mailto:vladyslavmalakho95@gmail.com)**

**Irina Gorbunova, Doctor Sc. (Pedagogy), Full Professor, Head of Computer Music Labs, Herzen University, Saint Petersburg, [gorbunovaib@rgpu.spb.ru](mailto:gorbunovaib@rgpu.spb.ru)**