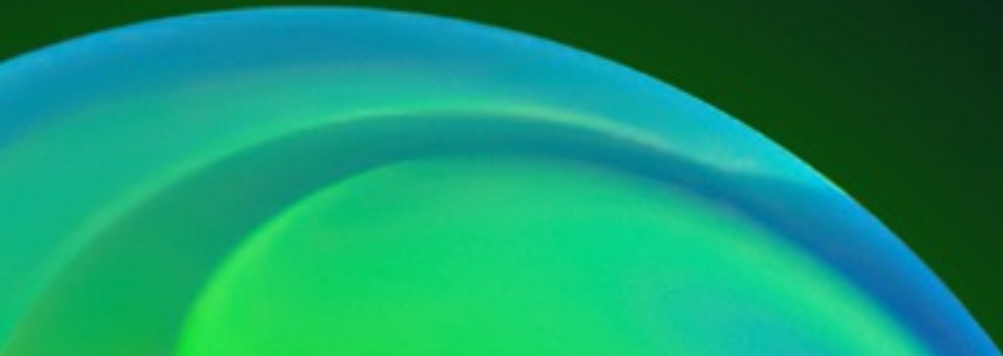


Интегральное скерцо:

как машины пишут музыку и поют

Дмитрий Ковалев, 25.04.2024



Какую из этих мелодий создал человек,
а какую — нейросеть?

1

2

Какую из этих мелодий создал человек,
а какую — нейросеть?



Какую из этих мелодий создал человек,
а какую — нейросеть?

1

2



Какую из этих мелодий создал человек,
а какую — нейросеть?

1

2

Какую из этих мелодий создал человек,
а какую — нейросеть?

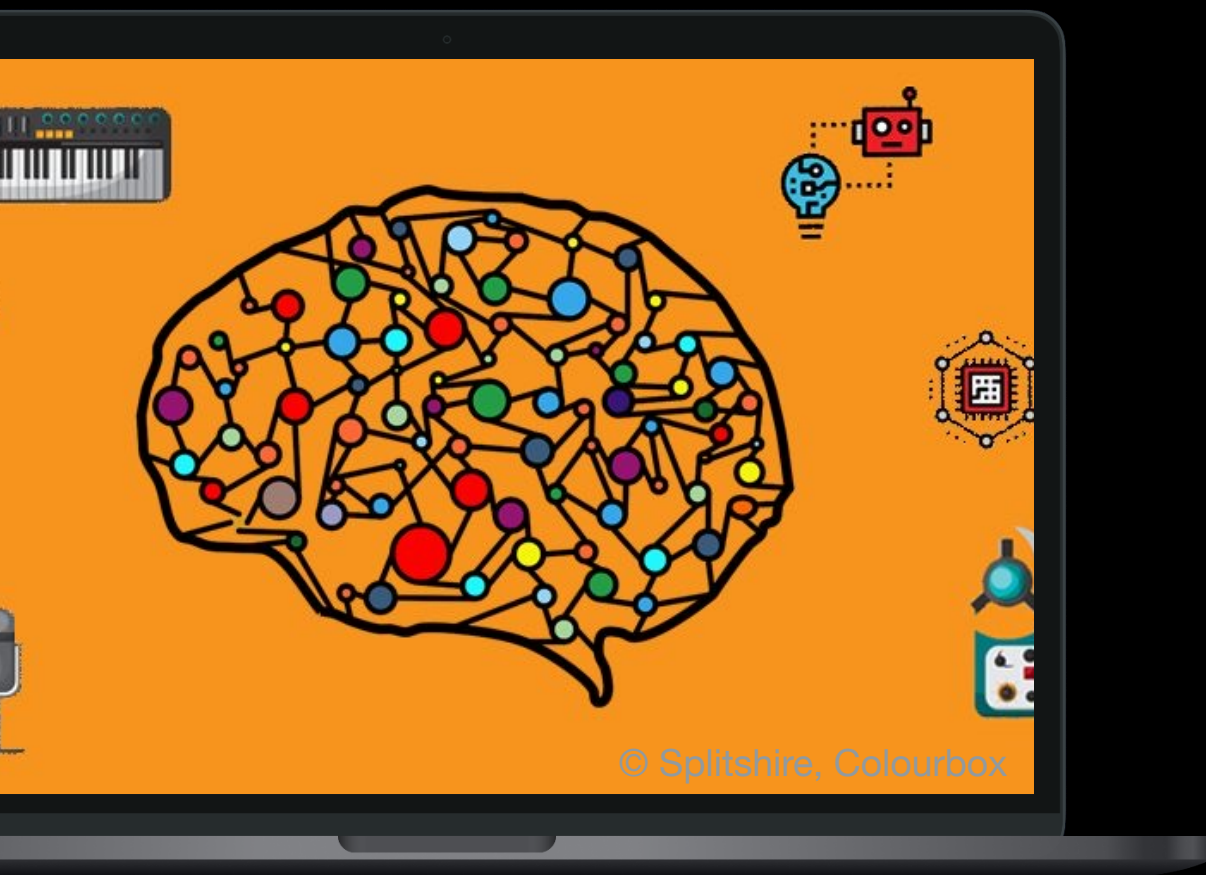
1

Человек

2

SymFormer

Музыка, ИИ и Музыкальный ИИ



Musica est exercitium arithmeticae occultum
nescientis se numerare animi
(G.W. Leibniz, 1712)

Искусственный интеллект - это наука и
технология создания интеллектуальных машин, в
особенности – интеллектуальных компьютерных
программ ... (John McCarthy, 1956)

... делающих нечто музыкальное на основе
музыкальных данных. (Keunwoo Choi, 2020)

Знаковые 2010 в музыкальном ИИ



Знаковые 2010 в музыкальном ИИ

2010, Tokyo

Hatsune Miku - Vocaloid (Yamaha),
бьющая все возможные рекорды продаж
и популярности дает первый в мире
сольный концерт с живыми музыкантами



Знаковые 2010 в музыкальном ИИ

2014, €50kk

Запущенный в 2005 по результатам диссертации MIT Media Lab сервис **EchoNest** приобретает **Spotify** для улучшения рекомендаций

2010, Tokyo

Hatsune Miku - Vocaloid (Yamaha), бьющая все возможные рекорды продаж и популярности дает первый в мире сольный концерт с живыми музыкантами



Знаковые 2010 в музыкальном ИИ

2014, €50kk

Запущенный в 2005 по результатам диссертации MIT Media Lab сервис **EchoNest** приобретает **Spotify** для улучшения рекомендаций

2010, Tokyo

Hatsune Miku - Vocaloid (Yamaha), бьющая все возможные рекорды продаж и популярности дает первый в мире сольный концерт с живыми музыкантами

2016, France

SACEM признает сервис **AIVA** как первого в мире виртуального композитора и закрепляет авторские права



Знаковые 2010 в музыкальном ИИ

2014, €50kk

Запущенный в 2005 по результатам диссертации MIT Media Lab сервис **EchoNest** приобретает **Spotify** для улучшения рекомендаций

2017, \$400kk+

Apple поглощает поисковик **Shazam**, запущенный в 1999 году

2010, Tokyo

Hatsune Miku - Vocaloid (Yamaha), бьющая все возможные рекорды продаж и популярности дает первый в мире сольный концерт с живыми музыкантами

2016, France

SACEM признает сервис **AIVA** как первого в мире виртуального композитора и закрепляет авторские права



Знаковые 2010 в музыкальном ИИ

2014, €50kk

Запущенный в 2005 по результатам диссертации MIT Media Lab сервис **EchoNest** приобретает **Spotify** для улучшения рекомендаций

2017, \$400kk+

Apple поглощает поисковик **Shazam**, запущенный в 1999 году

2010, Tokyo

Hatsume Miku - Vocaloid (Yamaha), бьющая все возможные рекорды продаж и популярности дает первый в мире сольный концерт с живыми музыкантами

2016, France

SACEM признает сервис **AIVA** как первого в мире виртуального композитора и закрепляет авторские права

2019, \$???kk

ByteDance (**TikTok**) покупает первый в мире публичный сервис для генерации музыки – **Jukedeck**, запущенный в 2012 году



С высоты птичьего полета

1

Автоматизированная
композиция сквозь призму
времени

2

Данные в
интеллектуальном
анализе музыки

3

Основные задачи,
строительные блоки

4

Поющие модели

5

Демо решений

6

Десерт =)

Начинаем раскопки

An On-Line Algorithm for Real-Time Accompaniment

Roger B. Dannenberg
Computer Science Department
Carnegie-Mellon University
Pittsburgh, PA 15213

Abstract

Real-time accompaniment solves most of the synchronization problems inherent in taped accompaniment; however, this new approach requires the computer to have the ability to follow the soloist. Three subproblems arise: detecting and processing input from the live performer, matching this input against a score of expected input, and generating the timing information necessary to control the generation of the accompaniment. It is expected that the live solo performance will contain mistakes or be imperfectly detected by the computer, so it is necessary to allow for performance mistakes when matching the actual solo against the score. An efficient dynamic programming algorithm for finding the best match between solo performance and the score is presented. In producing the accompaniment, it is necessary to generate a time reference that varies in speed according to the soloist. The notion of *virtual time* is proposed as a solution to this problem. Finally, experience with two computer systems that produce real-time accompaniment is summarized.

This research was sponsored by the Defense Advanced Research Projects Agency (DOD), ARPA Order No. 3597, monitored by the Air Force Avionics Laboratory Under Contract F33615-81-K-1539.

1. Introduction

Real-time digital music synthesis offers unprecedented flexibility to composers and performers of computer music. Applications have ranged from simple "player piano" systems to interactive improvisation involving both human and computer performers. An interesting possibility is that of a computer-generated accompaniment to a human soloist or ensemble. In this case, the computer is given a score containing parts for the soloist and for the corresponding accompaniment. Because the computer follows a human performer, most of the problems of synchronization between performers and computer-generated tapes is avoided. In fact, new artistic possibilities emerge because the accompaniment can respond to subtleties of the soloist's performance.

The problem of following a soloist can be divided into three subproblems. The first subproblem is to detect what the soloist is doing.¹ The second subproblem is to match the detected input against a score, which contains the expected input. This matching process must be tolerant of performance mistakes and of errors in the detection process. The third subproblem is to produce an accompaniment that follows the soloist. Techniques are needed here to perform the accompaniment at the appropriate rate between detected solo events, and to recover smoothly when it is determined that the accompaniment and soloist have lost their synchronization.

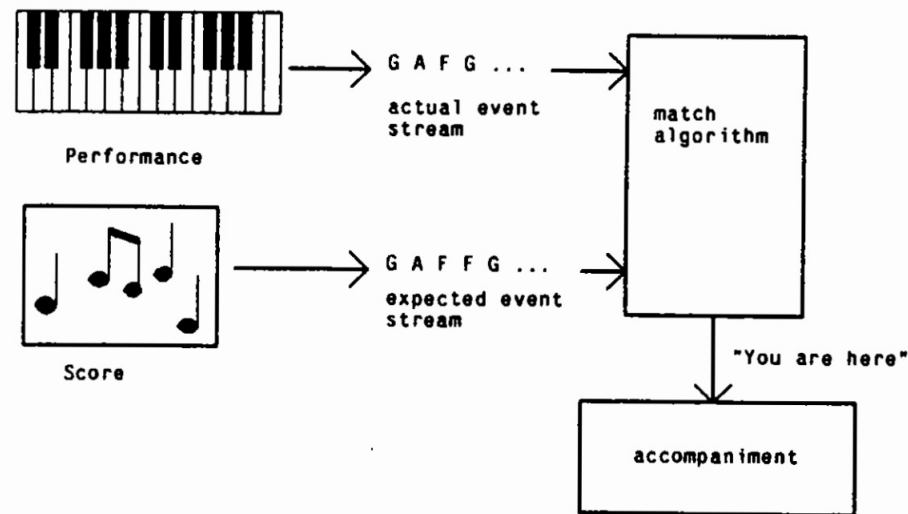


Figure 2-1: The actual event stream from the performance is compared to the expected event stream from the score to determine the present location.

Начинаем раскопки

An On-Line Algorithm for Real-Time Accompaniment

Roger B. Dannenberg
Computer Science Department
Carnegie-Mellon University
Pittsburgh, PA 15213

Abstract

Real-time accompaniment solves most of the synchronization problems inherent in taped accompaniment; however, this new approach requires the computer to have the ability to follow the soloist. Three subproblems arise: detecting and processing input from the live performer, matching this input against a score of expected input, and generating the timing information necessary to control the generation of the accompaniment. It is expected that the live solo performance will contain mistakes or be imperfectly detected by the computer, so it is necessary to allow for performance mistakes when matching the actual solo against the score. An efficient dynamic programming algorithm for finding the best match between solo performance and the score is presented. In producing the accompaniment, it is necessary to generate a time reference that varies in speed according to the soloist. The notion of *virtual time* is proposed as a solution to this problem. Finally, experience with two computer systems that produce real-time accompaniment is summarized.

This research was sponsored by the Defense Advanced Research Projects Agency (DOD), ARPA Order No. 3597, monitored by the Air Force Avionics Laboratory Under Contract F33615-81-K-1539.

1. Introduction

Real-time digital music flexibility to compose music. Applications to "piano" systems to integrate both human and computer possibility is that accompaniment to a human case, the computer is the soloist and for the computer. Because the computer is the soloist and for the computer, the problems of synchronization and computer-generated artistic possibilities can respond to subtle

The problem of following three subproblems. The first is what the soloist is doing and matching the detected input to the expected input. The second is to be tolerant of performance detection process. The third is the accompaniment that is needed here to perform at an appropriate rate between recovering smoothly when the accompaniment and synchronization.

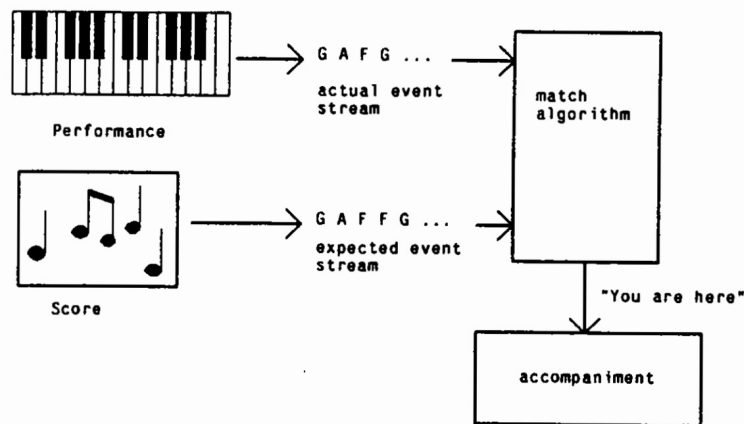


Figure 2-1: The actual event stream from the performance is compared to the expected event stream from the score to determine the present location.

References

1. P. J. Bloom. Use of Dynamic Programming for Automatic Synchronization of Two Similar Speech Signals. Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, 1984, pp. 2.6.1-2.6.4.
2. W. Buxton, W. Reeves, G. Fedorkow, K. C. Smith, and R. Baecker. "A Microcomputer-based Conducting System." *Computer Music Journal* 4, 4 (Spring 1980), 8-21.
3. Chris Chafe, Bernard Mont-Reynaud, and Loren Rush. "Toward an Intelligent Editor of Digital Audio: Recognition of Musical Constructs." *Computer Music Journal* 6, 1 (Spring 1982), 30-41.
4. Jane Clendinning and Paul E. Dworak. Computer Pitch Recognition: A New Approach. 1983 ICMC Proceedings, Computer Music Association, 1983.
5. Scott Foster, W. Andrew Schloss, A. Joseph Rockmore. "Toward an Intelligent Editor of Digital Audio: Signal Processing Methods." *Computer Music Journal* 6, 1 (Spring 1982), 42-51.
6. Steven M. Haflich and Mark A. Burns. Following a Conductor: The Engineering of an Input Device. 1983 ICMC Proceedings, Computer Music Association, 1983.
7. Max V. Mathews and Curtis Abbot. "The Sequential Drum." *Computer Music Journal* 4, 4 (Winter 1980), 45-59.
8. John Snell. "The Lucasfilm Real-Time Console for Recording Studios and Performance of Computer Music." *Computer Music Journal* 6, 3 (Fall 1982), 33-45.
9. diffreg.c. Source code for the 4.1BSD Berkeley Unix™ *diff* program.
10. A. Waibel, N. Krishnan, R. Reddy. Minimizing Computational Cost for Dynamic Programming Algorithms. Tech. Rept. CMU-CS-81-124, Carnegie-Mellon University Department of Computer Science, June, 1981.
11. A. Waibel and B. Yegnanarayana. Comparative Study of Nonlinear Time Warping Techniques in Isolated Word Speech Recognition Systems. Tech. Rept. CMU-CS-81-125, Carnegie-Mellon University Department of Computer Science, June, 1981.

Musikalisches Würfelspiel, W.A.Mozart, 1787

INSTRUCTION

To compose without the least knowledge of Music,
German Walzer or Schleifer, by throwing a
certain Number with two Dice.

- 1., The letters A—H, placed at the head of the 8 Columns of the Number-Tables show the 8. times of each part of the Walzer. VIZ. A, the first, B, the second, C, the third, &c. and the numbers in the Column under the letters, show the number of the time in the notes.
- 2., the numbers from 2 to 12 show the sum of the number that can be thrown.
- 3., For instance, in throwing for the first time of the first part of the Walzer, with two dice, the number 6, one looks next to that number in the Column A, for the 1st time in the notes. This time is written down, and makes the beginning of the Walzer. — For the second time, for instance, the number 9, being thrown, turn to the same table Column H, and the number 8 $\frac{1}{2}$ shall be found. This time is put next to the first, & one continues, in this manner, till the dice shall be thrown all the eight times, when likewise the first part of the Walzer shall be finish'd. The sign of repetition is further plac'd & the second part begun, & in case a still longer Walzer be desired, the beginning is again in the same manner, & one continues as long as one pleases.

6.

129. 130. 131. 132. 133. 134. 135. 136.

137. 138. 139. 140. 141. 142. 143. 144.

145. 146. 147. 148. 149. 150. 151. 152.

153. 154. 155. 156. 157. 158. 159. 160.

161. 162. 163. 164. 165. 166. 167. 168.

169. 170. 171. 172. 173. 174. 175. 176.

Illinois Automatic Computer & ILLIAC Suite, 1957-1960



Illinois Automatic Computer & ILLIAC Suite, 1957-1960

Четыре части = четыре эксперимента

1. Генерация мелодии по строгим правилам
2. Ансамблирование на 4 голоса правилами
3. Рандомизация длительностей и ритмов, разрешили «хроматизмы»
4. Марковские цепи и отказ от правил

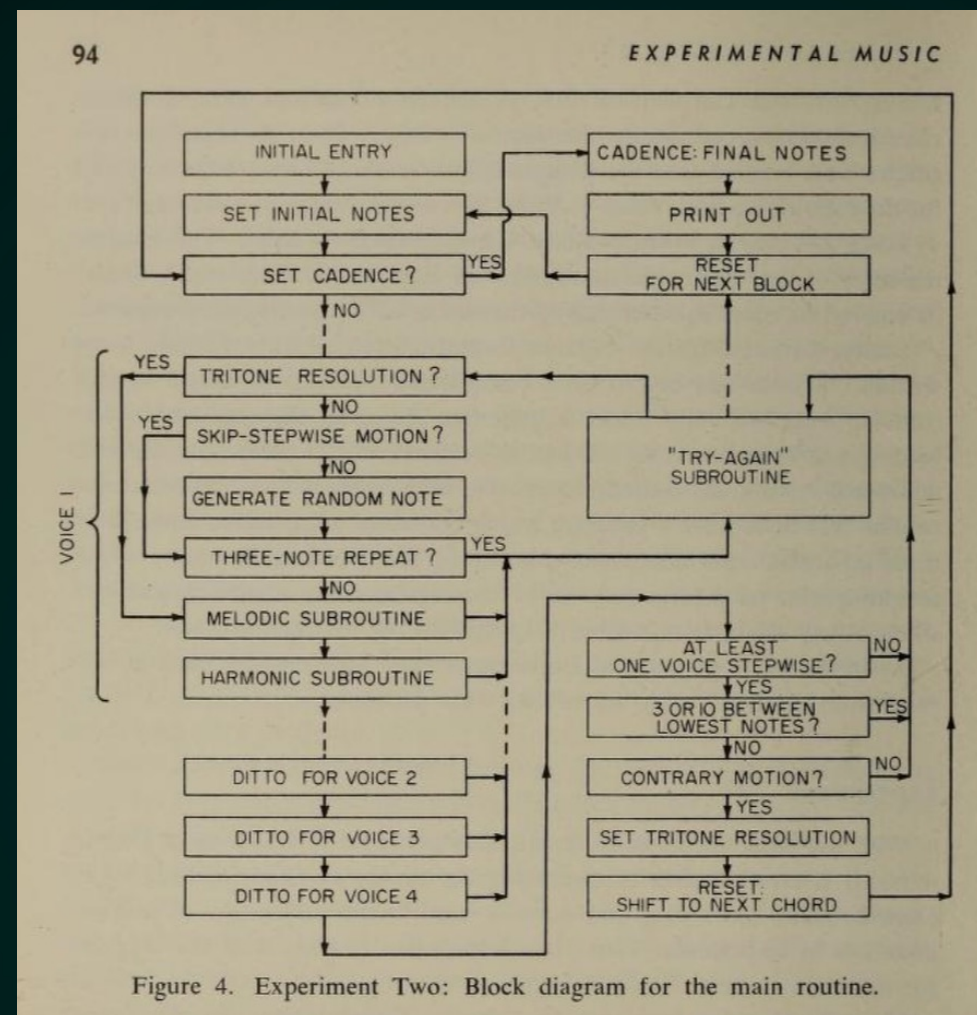


Figure 4. Experiment Two: Block diagram for the main routine.

Illinois Automatic Computer & ILLIAC Suite, 1957-1960

Четыре части = четыре эксперимента

1. Генерация мелодии по строгим правилам
2. Ансамблирование на 4 голоса правилами
3. Рандомизация длительностей и ритмов, разрешили «хроматизмы»
4. Марковские цепи и отказ от правил

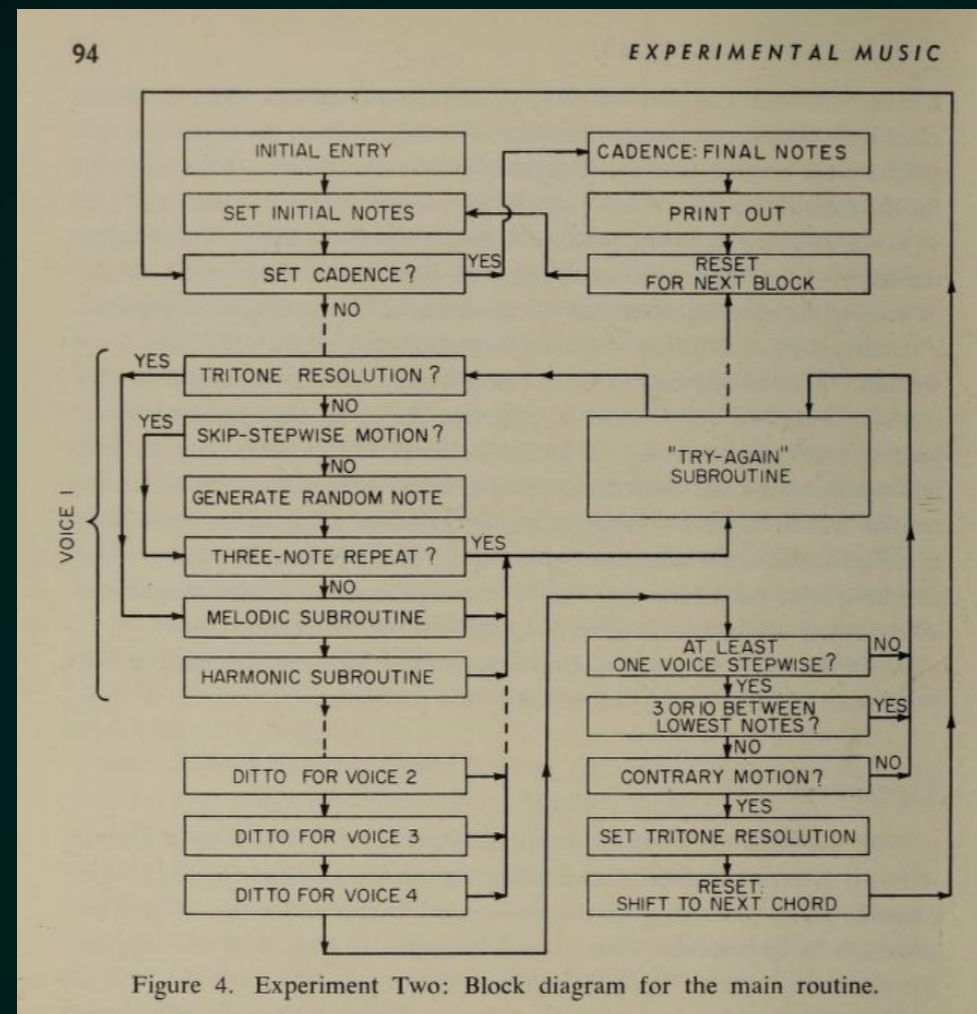


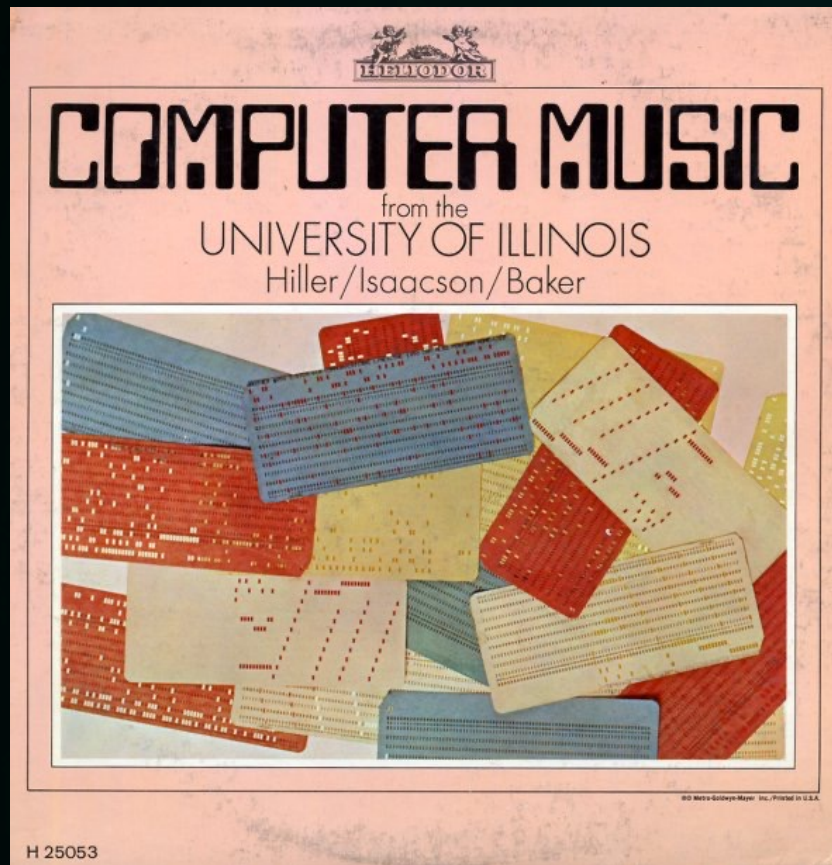
Figure 4. Experiment Two: Block diagram for the main routine.

Illinois Automatic Computer & ILLIAC Suite, 1957-1960

Lejaren Hiller - *Illiad Suite for String Quartet* (1956)

First experiment: presto, andante, allegro

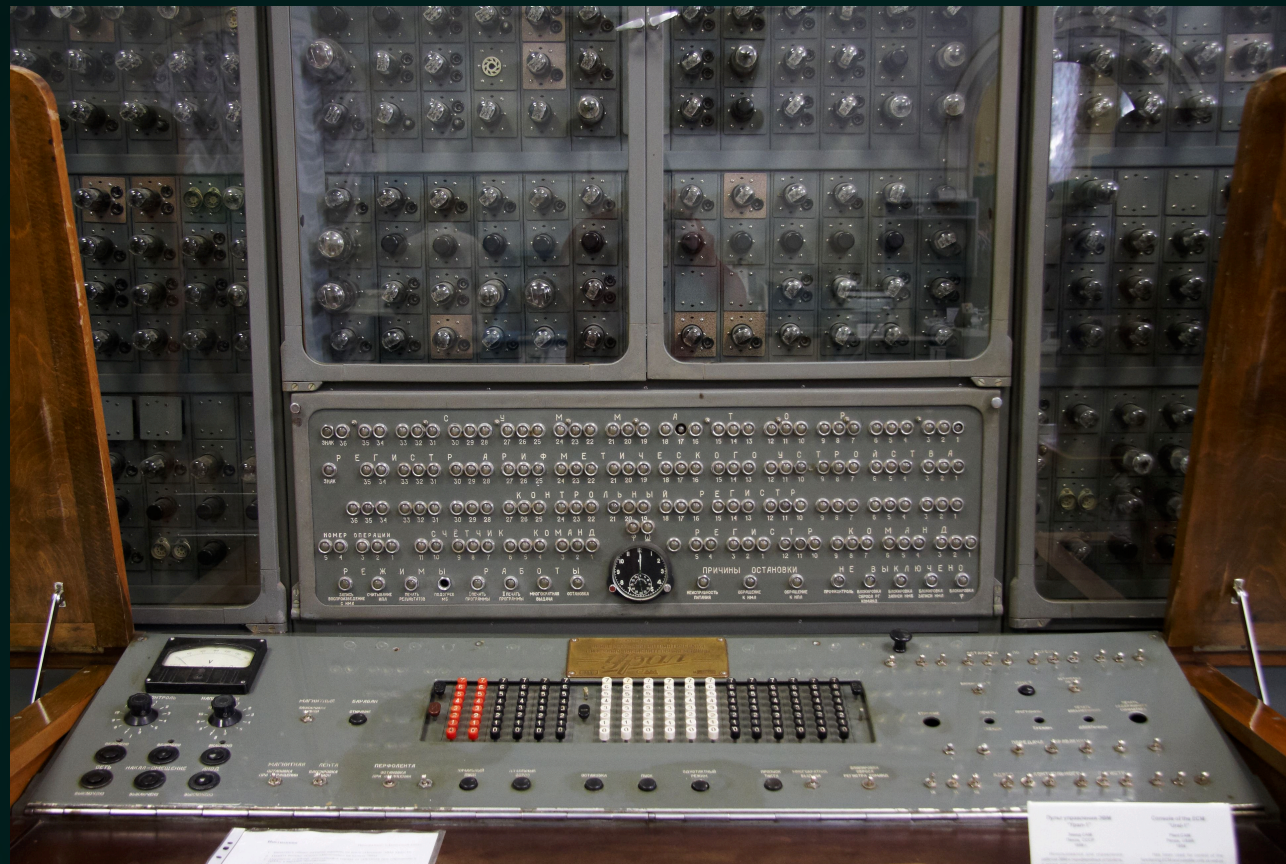
Illinois Automatic Computer & ILLIAC Suite, 1957-1960



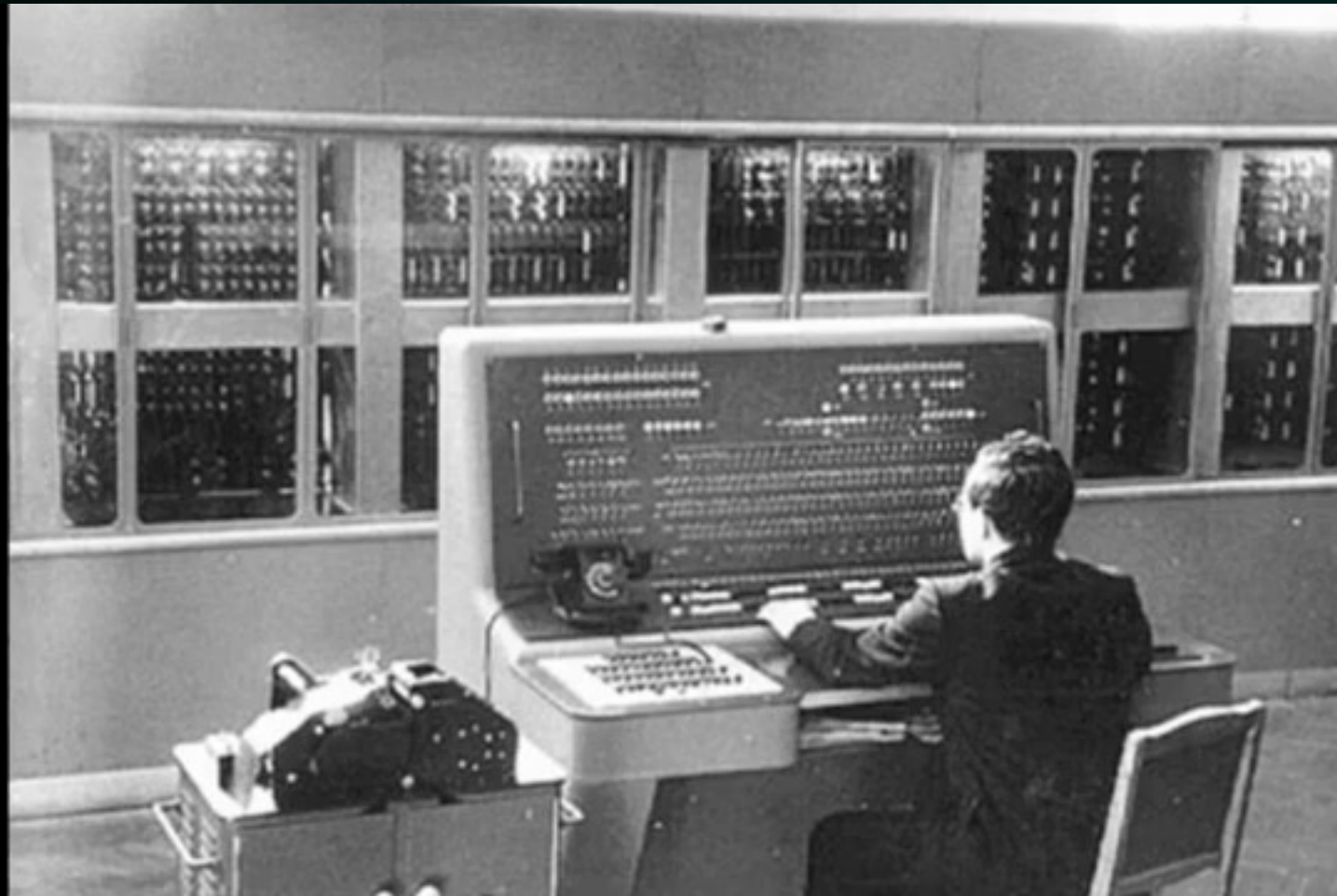
Рудольф Зарипов и Урал-1 , 1959-1960

«Об алгоритмическом описании процесса сочинения музыки»

- Одноголосная композиция
- Алгоритмические правила
- Нота – пятизначное число №_L_F0



Рудольф Зарипов и Урал-1 , 1959-1960

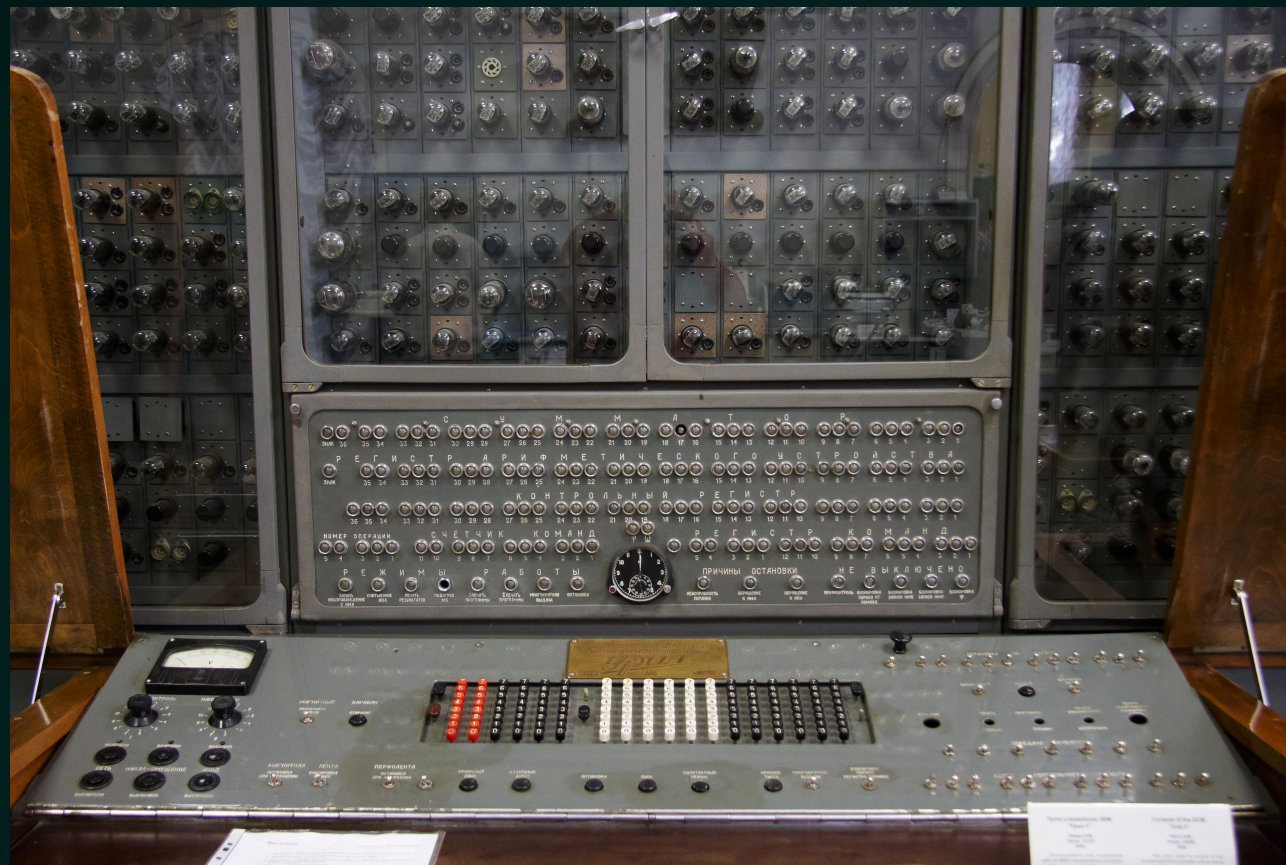


Рудольф Зарипов и Урал-1 , 1959-1960

«Об алгоритмическом описании процесса сочинения музыки»

- Одноголосная композиция
- Алгоритмические правила
- Нота – пятизначное число №_L_F0

1. Генерация структуры
2. Выбор ритмического рисунка частей
3. Рандомизация нот по правилам
4. Проверка результата



Современное положение дел

Аудио

-
- | | |
|-------|---|
| Аудио | <ul style="list-style-type: none">- Source (Instrument) Separation- Auto-mixing- Style transfer ... |
|-------|---|
-
-
-

Современное положение дел

	Аудио	Символы
Аудио	<ul style="list-style-type: none">- Source (Instrument) Separation- Auto-mixing...	<ul style="list-style-type: none">- Waveform synthesis- Performance generation...
Символы	<ul style="list-style-type: none">- Automatic Music Transcription- Alignment...	<ul style="list-style-type: none">- Automatic composition- Symbolic music generation...

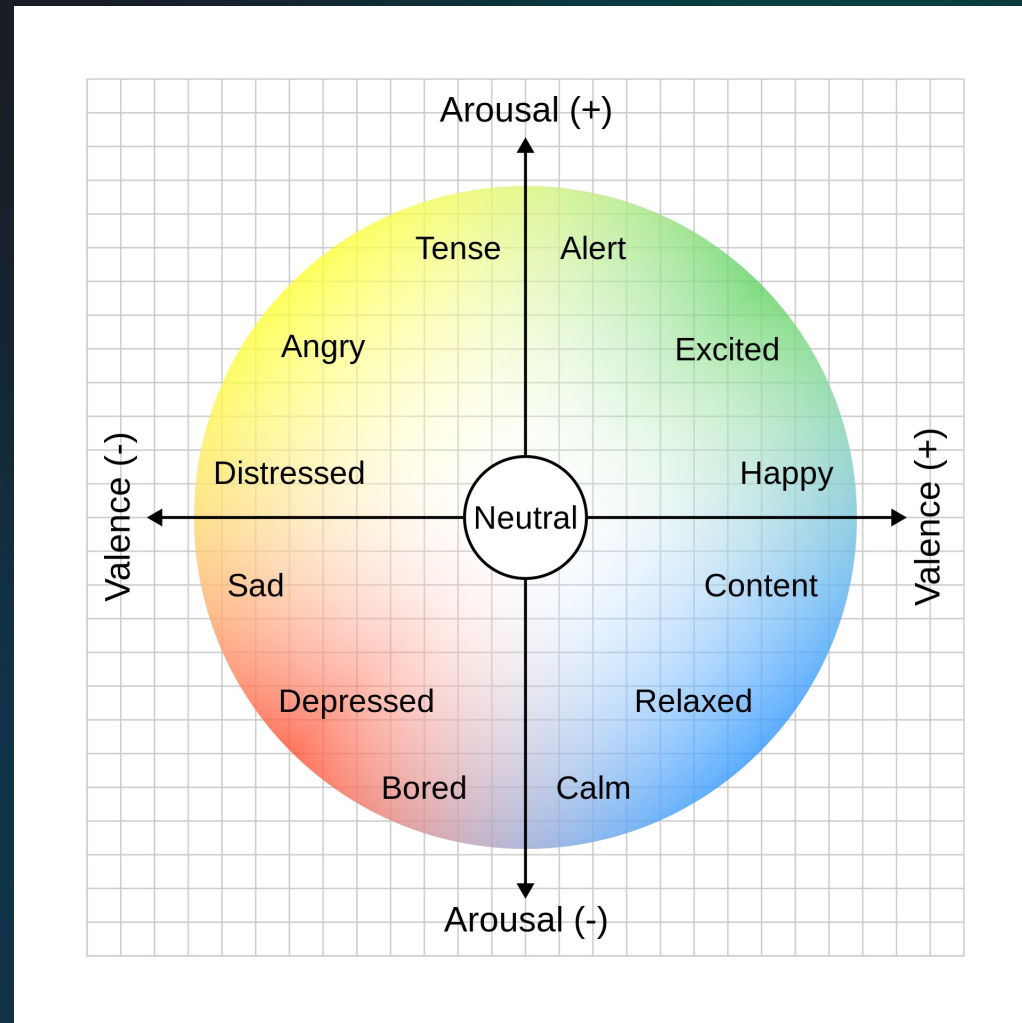
Современное положение дел

	Аудио	Символы	Метаданные и тексты
Аудио	<ul style="list-style-type: none">- Source (Instrument) Separation- Auto-mixing...	<ul style="list-style-type: none">- Waveform synthesis- Performance generation...	<ul style="list-style-type: none">- Singing Voice Generation- Text-to-music- Controllable Music generation...
Символы	<ul style="list-style-type: none">- Automatic Music Transcription- Alignment...	<ul style="list-style-type: none">- Automatic composition- Symbolic music generation...	
Мета	<ul style="list-style-type: none">- Genre classification / music auto-tagging- Music Captioning- Music Question Answering...		<ul style="list-style-type: none">- Music pseudo-captioning...

Современное положение дел

	Аудио	Символы	Метаданные и тексты
Аудио	<ul style="list-style-type: none">- Source (Instrument) Separation- Auto-mixing...	<ul style="list-style-type: none">- Waveform synthesis- Performance generation...	<ul style="list-style-type: none">- Singing Voice Generation- Text-to-music- Controllable Music generation...
Символы	<ul style="list-style-type: none">- Automatic Music Transcription- Alignment...	<ul style="list-style-type: none">- Automatic composition- Symbolic music generation...	
Мета	<ul style="list-style-type: none">- Genre classification / music auto-tagging- Music Captioning- Music Question Answering...		<ul style="list-style-type: none">- Music pseudo-captioning...

Угадай эмоцию!

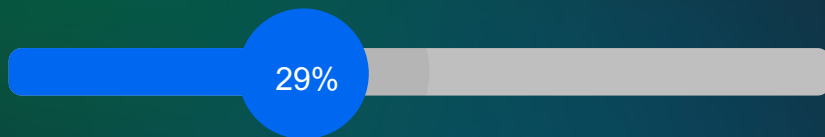


Угадай эмоцию!

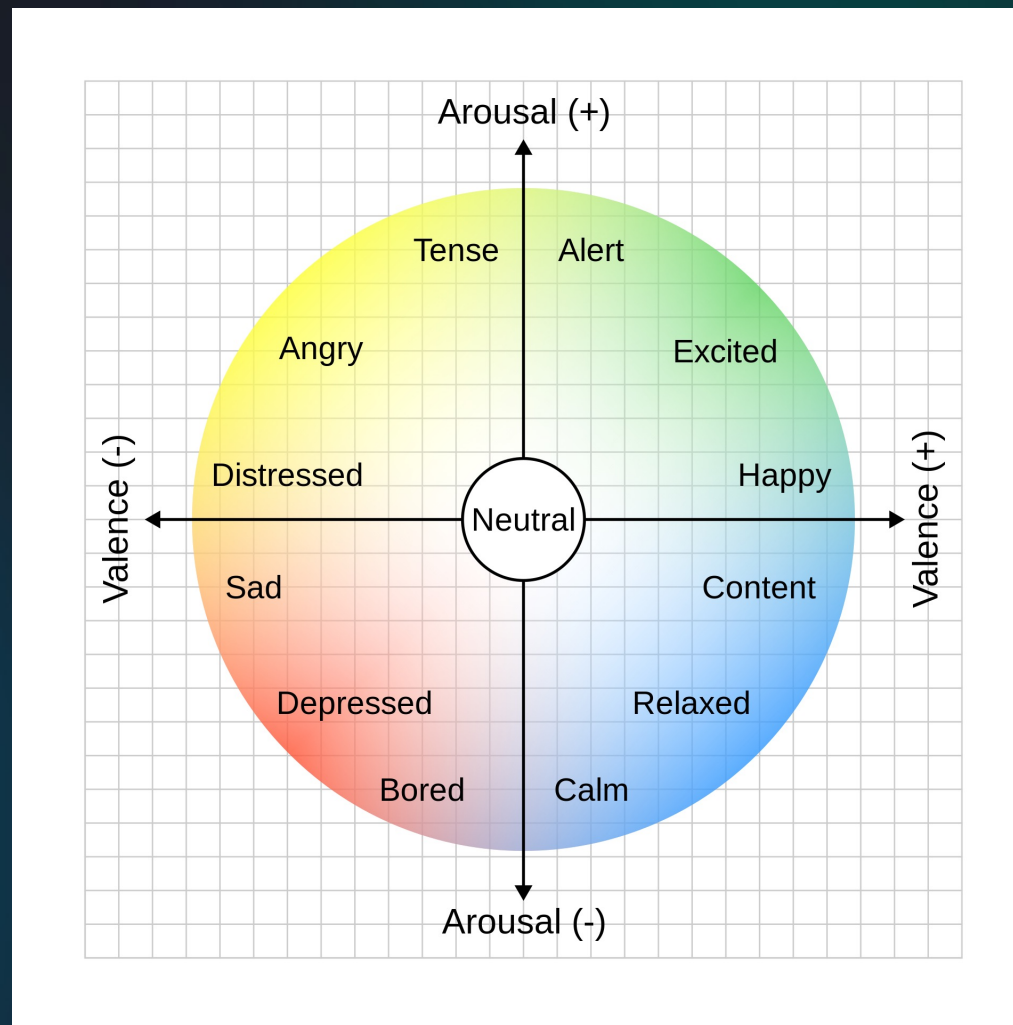
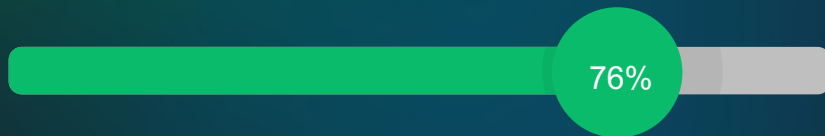


EMOPIA, 2021

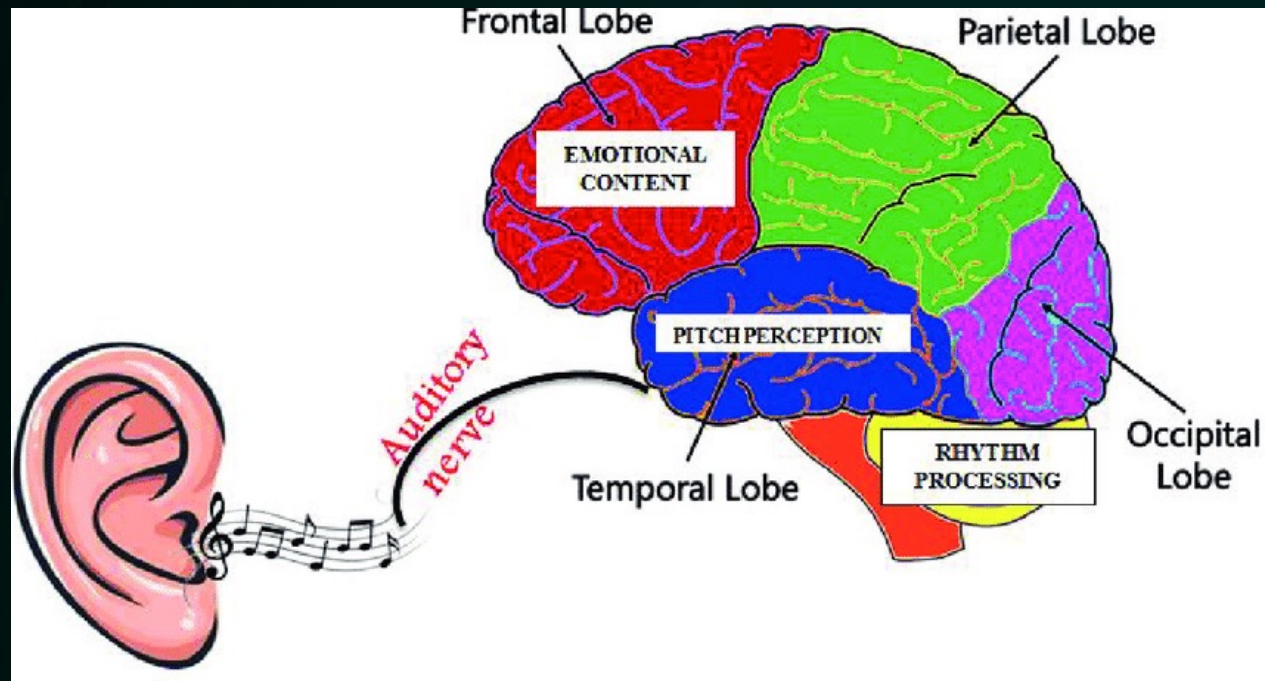
Valence



Arousal

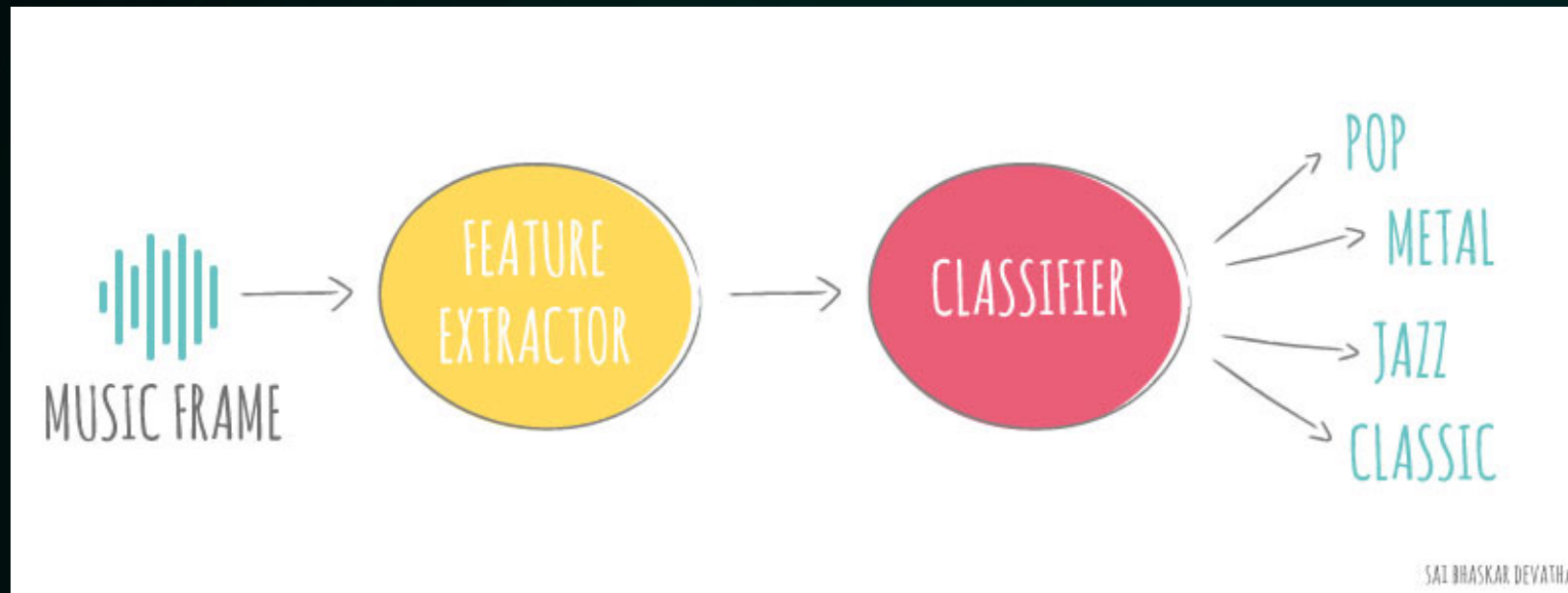


Классификация музыки

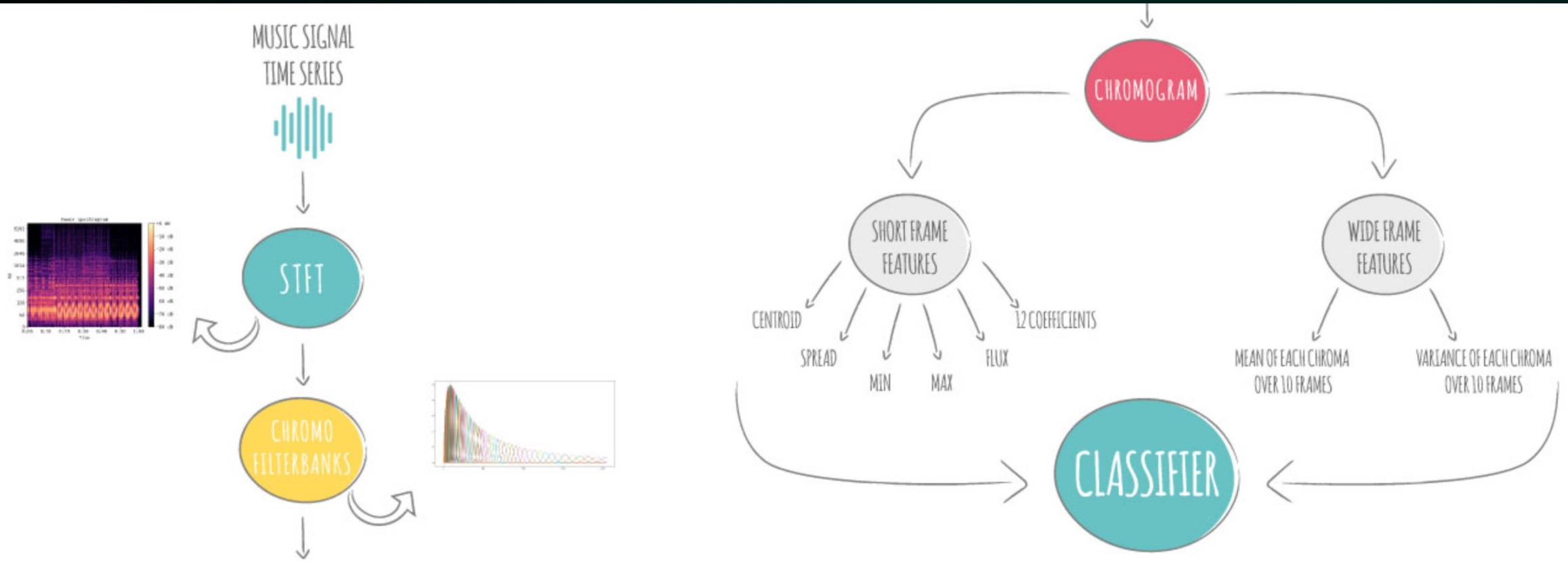


Напряженная!

Классификация музыки



Классификация музыки



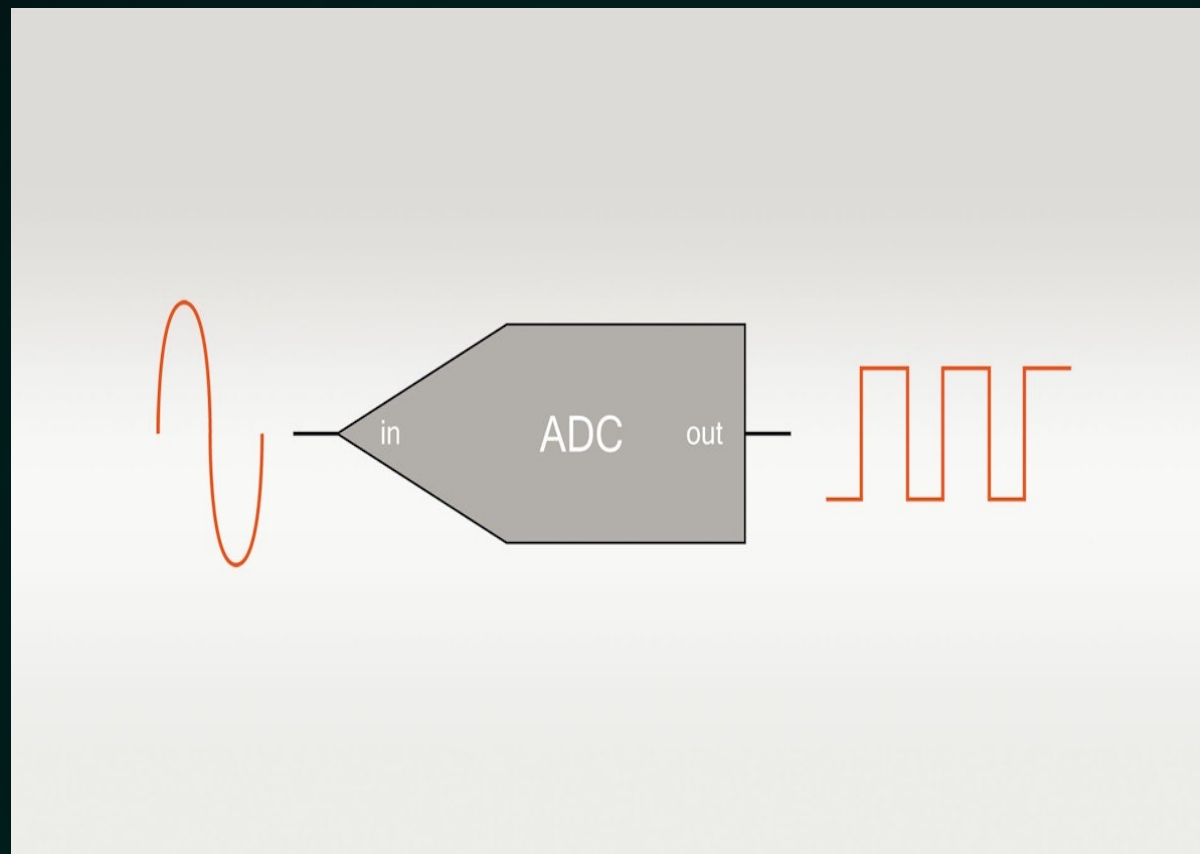
Картинка со звуком



"Painty the Pirate" Painted by Calvin Liang 4-99.

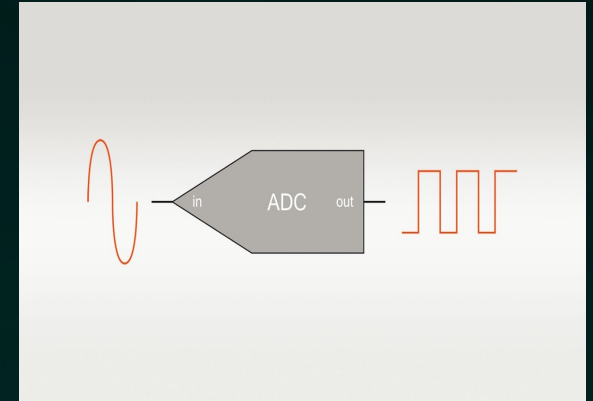
Акустическая волна. АЦП

- 8 000 Гц — телефон, речь
- 22 050 Гц — радио
- ...



Акустическая волна. АЦП

- 8 000 Гц — телефон, речь
- 22 050 Гц — радио
- 44 100 Гц — Audio CD



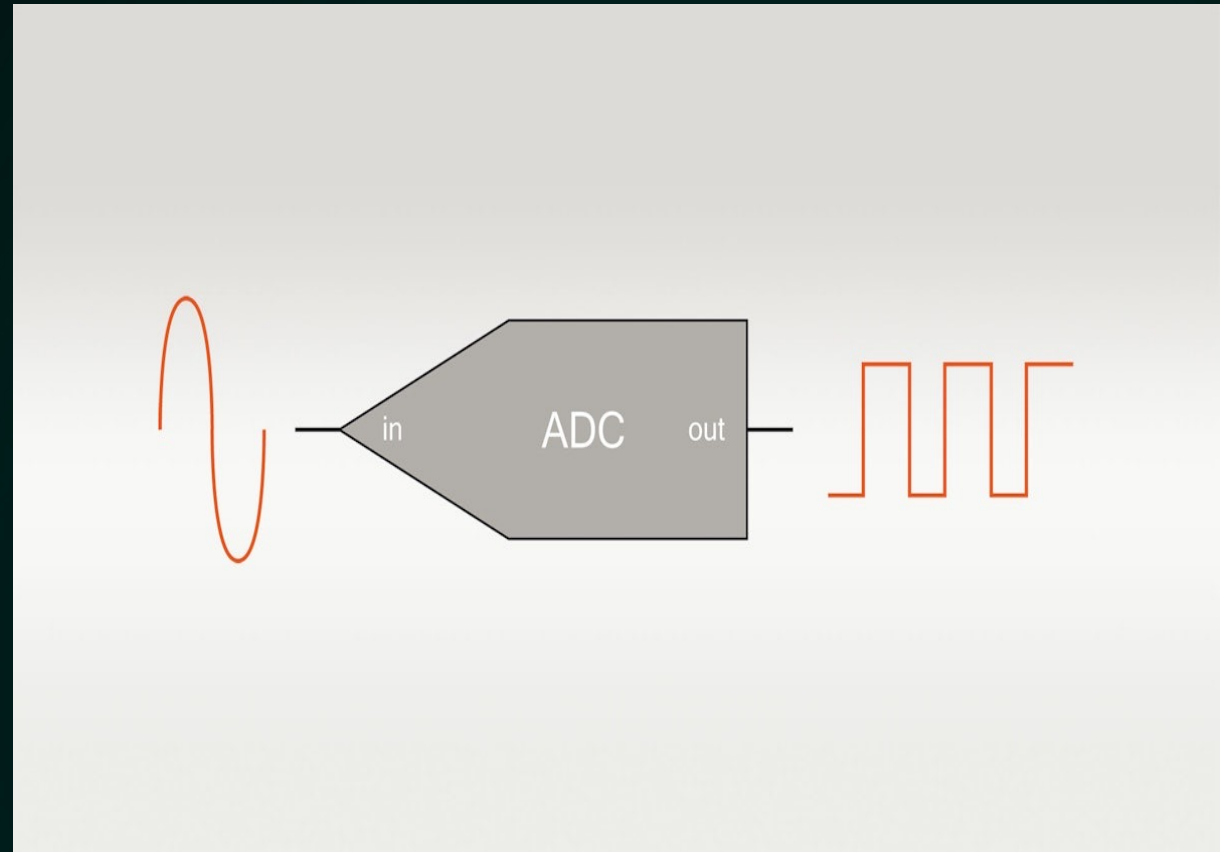
По теореме Котельникова 44кГц
покрывает весь слышимый
человеческом звук
20 Гц — 20 КГц

$$x(t) = \sum_{k=-\infty}^{\infty} x(k\Delta) \operatorname{sinc} \left[\frac{\pi}{\Delta} (t - k\Delta) \right]$$

$$0 < \Delta \leq \frac{1}{2f_c}$$

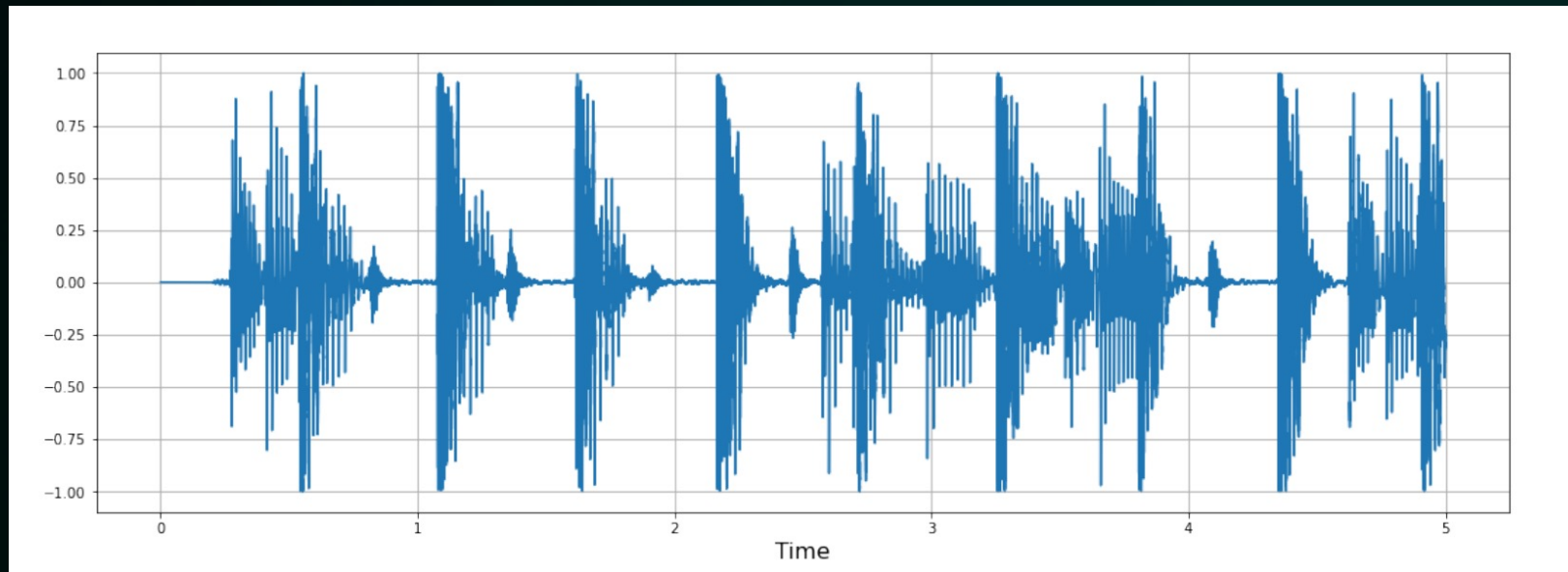
Акустическая волна. АЦП

- 8 000 Гц — телефон, речь
- 22 050 Гц — радио
- 44 100 Гц — Audio CD
- 96 000 Гц — DVD-Audio
- 2 822 400 Гц — Super Audio CD



Дискретная акустическая волна

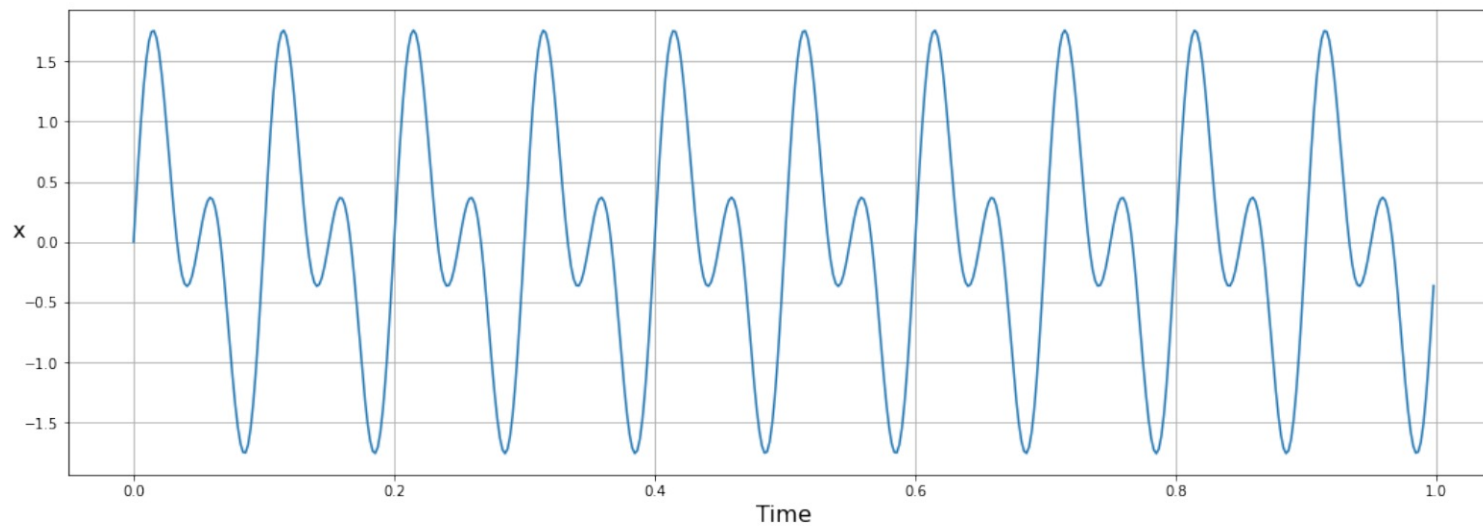
$$x = (x(0), \dots, x(N - 1)), x(n) \in [-1, 1], n \in [0, N - 1]$$



Первые 5 секунд "Another One Bites the Dust" группы Queen

Дискретное преобразование Фурье, DFT

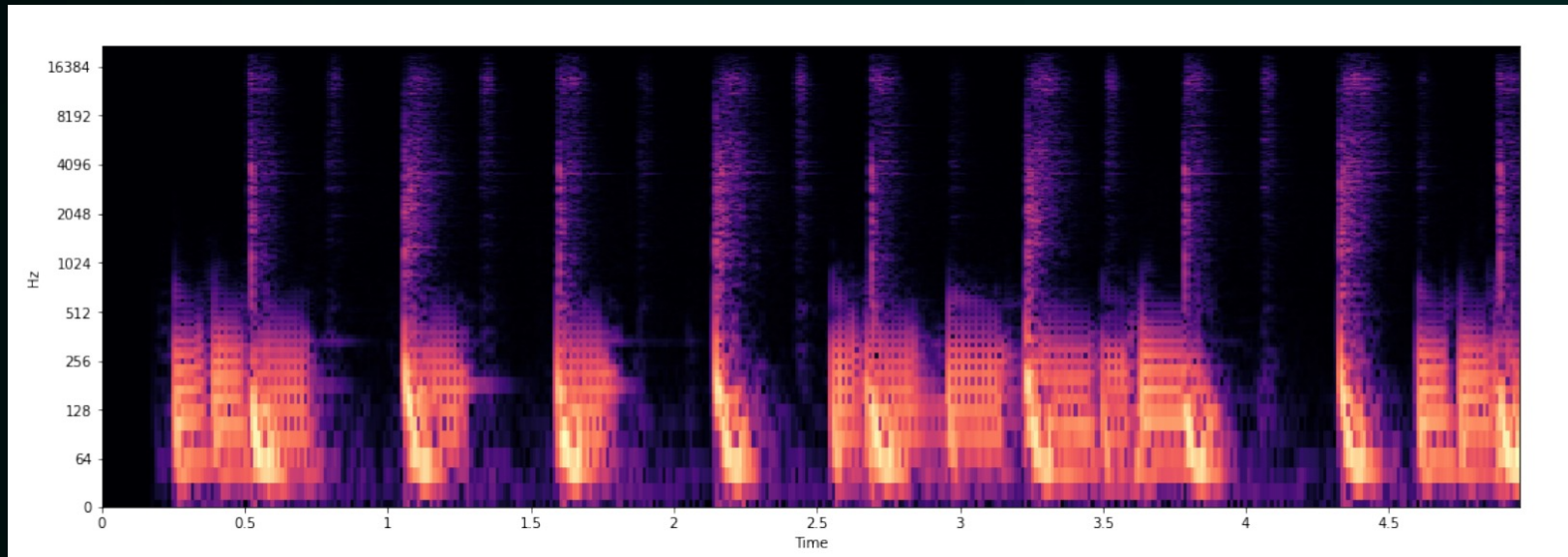
$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N} kn} = \sum_{n=0}^{N-1} x_n \left(\cos(2\pi kn/N) - i \cdot \sin(2\pi kn/N) \right), \quad k = 0, \dots, N - 1.$$



Сигнал $x = \sin(10 \cdot 2\pi t) + \sin(20 \cdot 2\pi t)$, $t \in [0, 1]$

Оконное преобразование Фурье, STFT. Спектрограмма

$$S_x(i, k) = |STFT_x(i, k)|, \quad STFT_x(i, k) = DFT_{x_i^w}(k), \quad k \in [0, W - 1], \quad i \in [0, I - 1]$$



Первые 5 секунд "Another One Bites the Dust" группы Queen

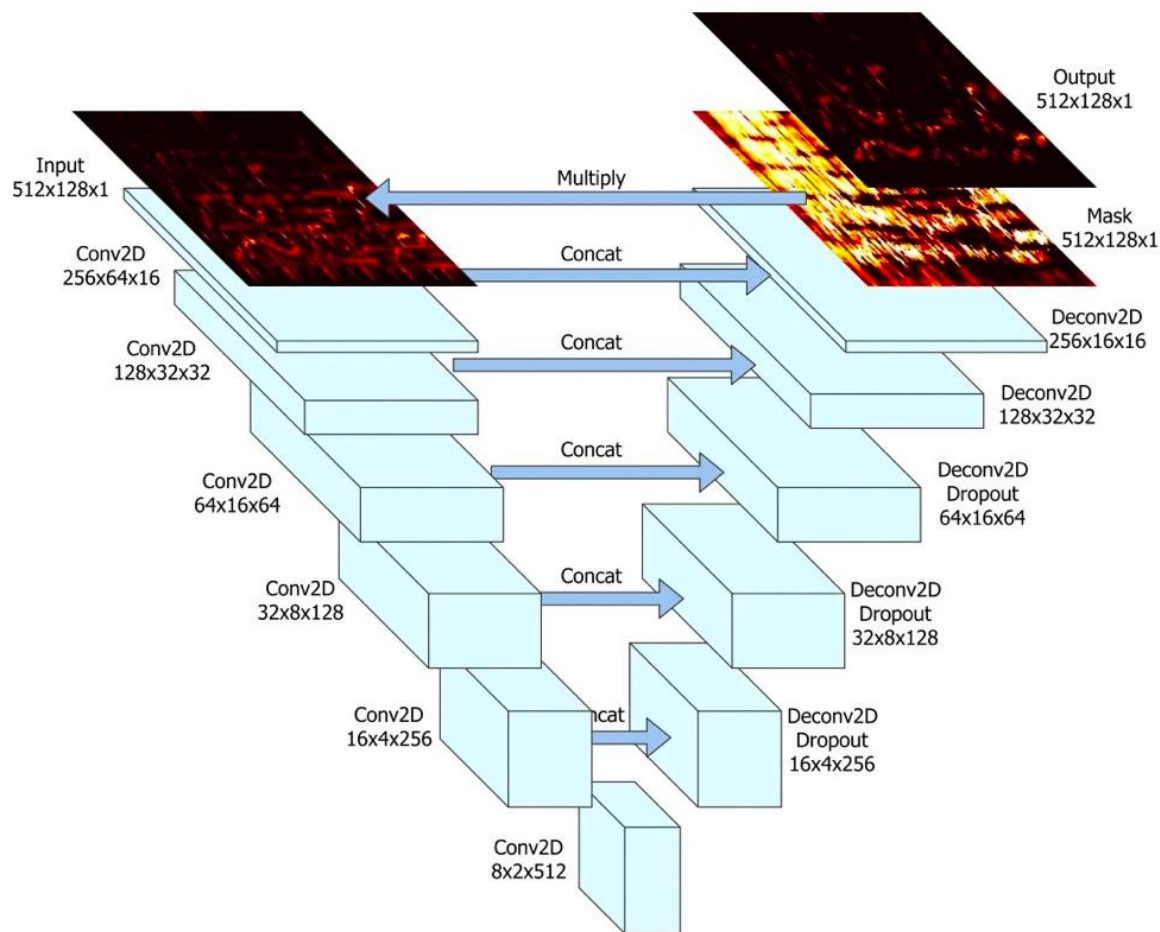


Идея!

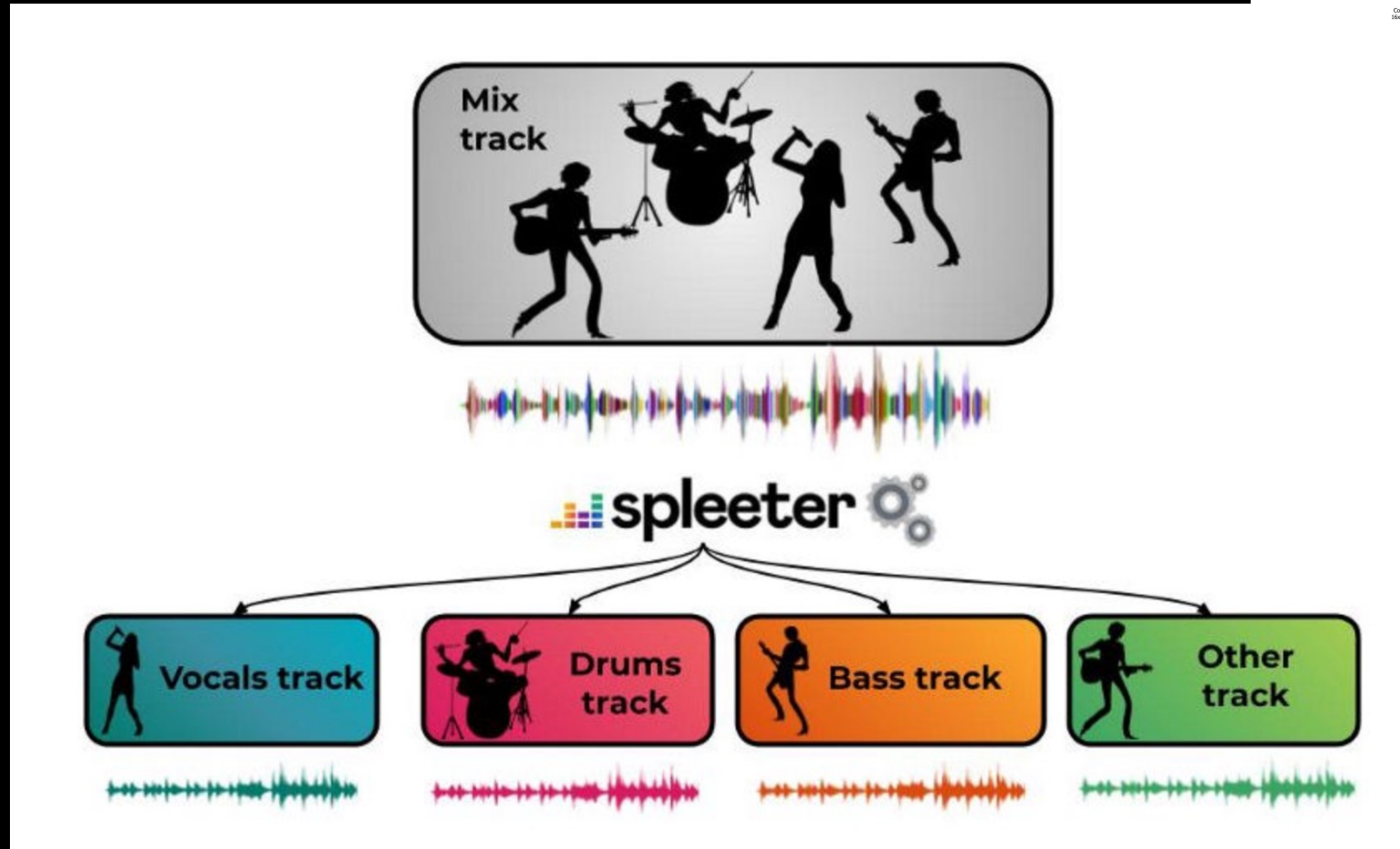
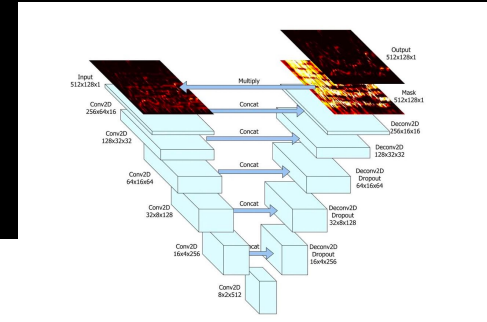
Музыка как картинка!

Пилите спектрограммы, они золотые!

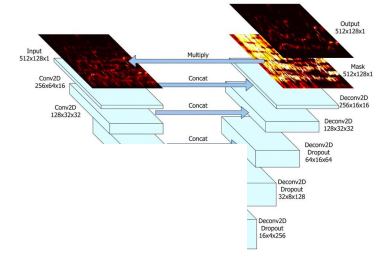
U-net



Music Source Separation. Spleeter



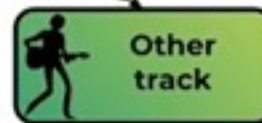
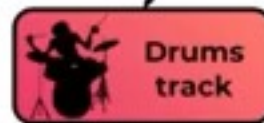
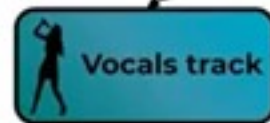
Music Source Separation. Spleeter



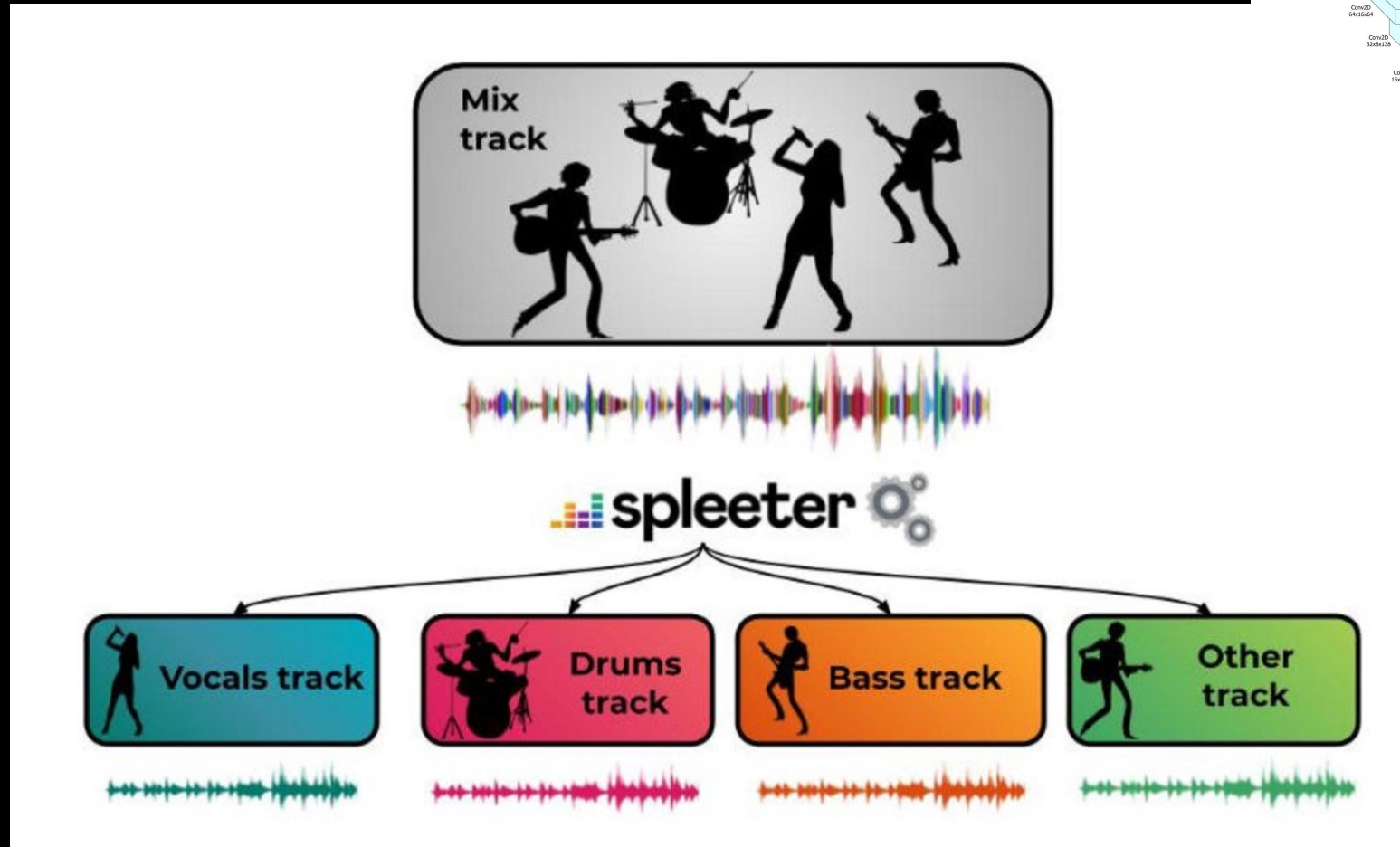
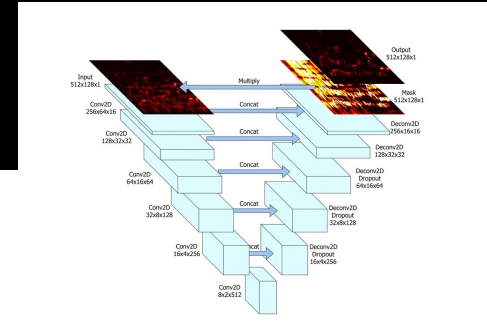
Spleeter - tool
for music separation [Demo]



spleeter



Music Source Separation. Spleeter



Компоненты звука. Синтезатор



Компоненты звука. Синтезатор



Клавиши
ТОН

Компоненты звука. Синтезатор



Клавиши
ТОН

Ручки
Тембры

Компоненты звука. Синтезатор



Клавиши
ТОН

Ручки
Тембры

Громкость

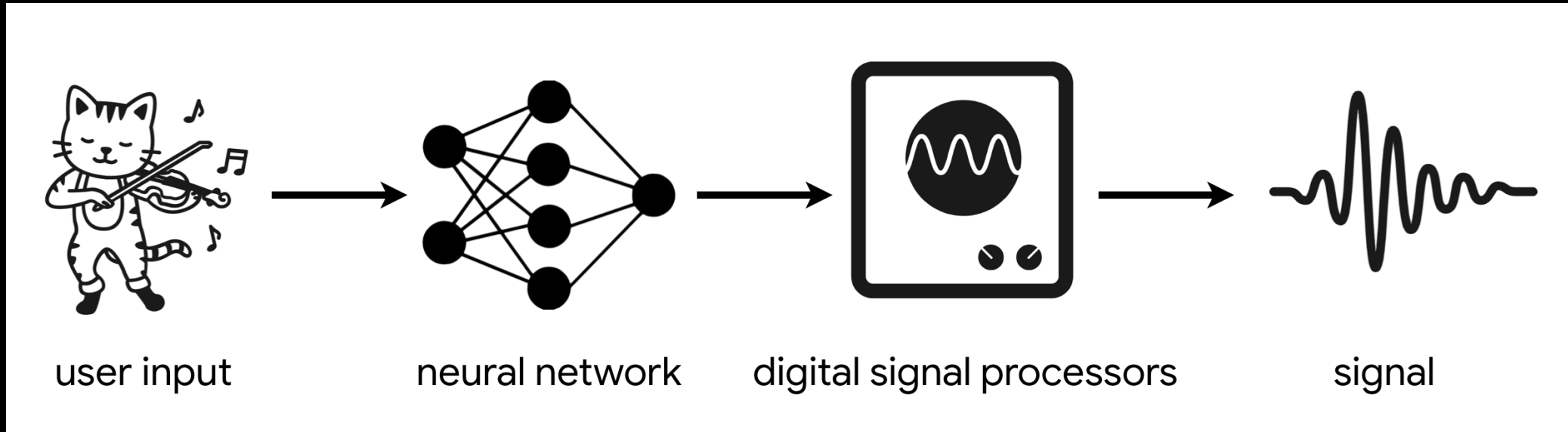
Компоненты звука. Синтезатор



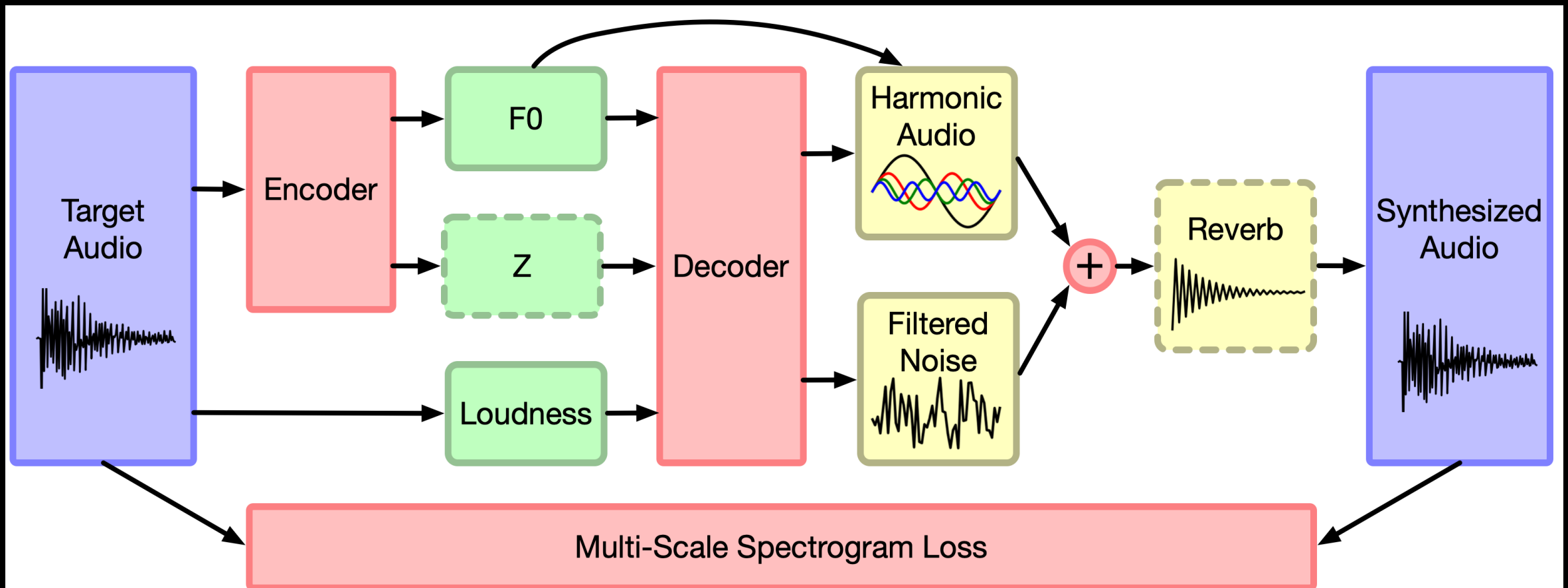
Синтезатор. Декодировщик



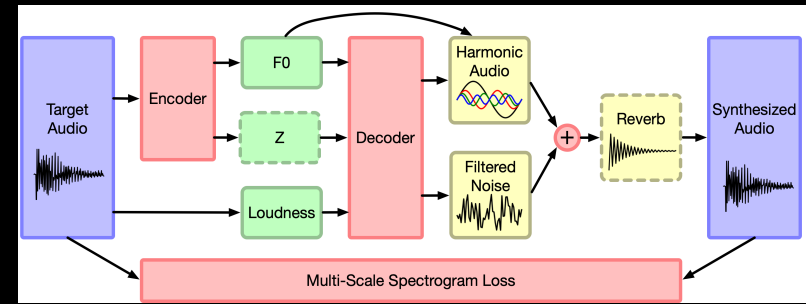
Music Audio Synthesis. DDSP



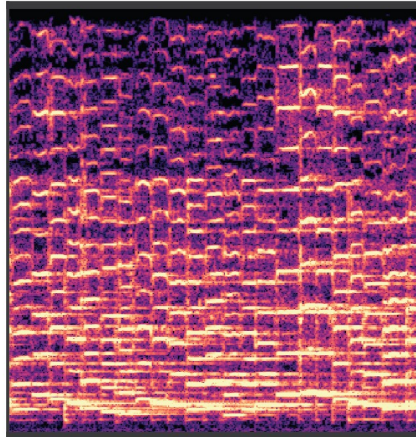
Music Audio Synthesis. DDSP



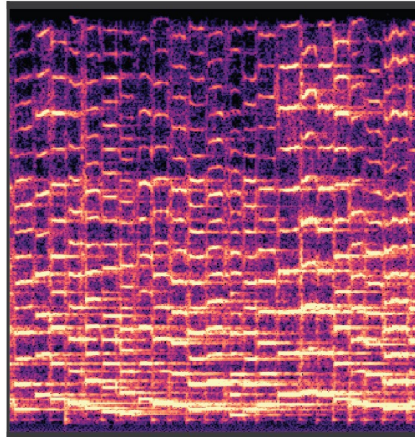
Music Audio Synthesis. DDSP



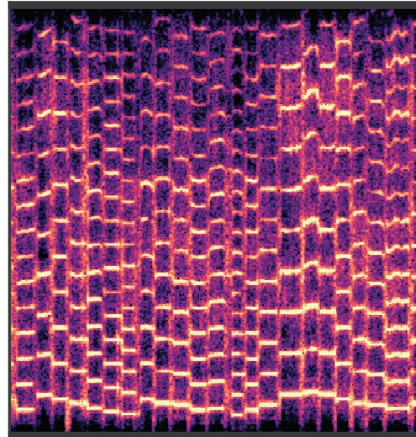
Original Audio



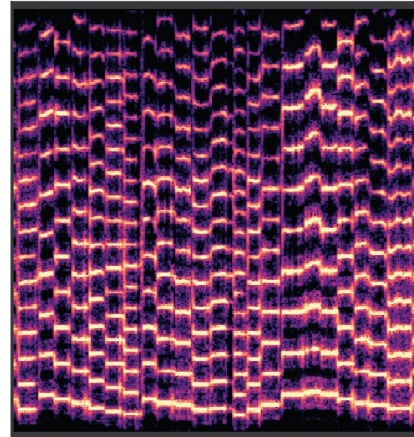
Resynthesis



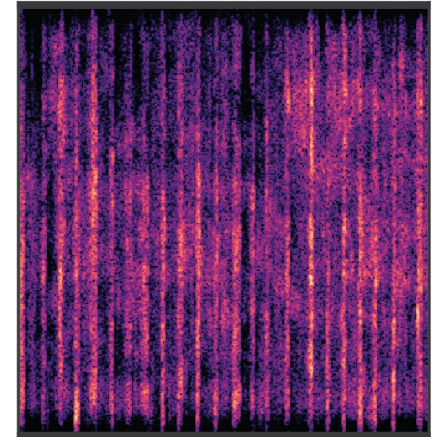
Reverb
Removed



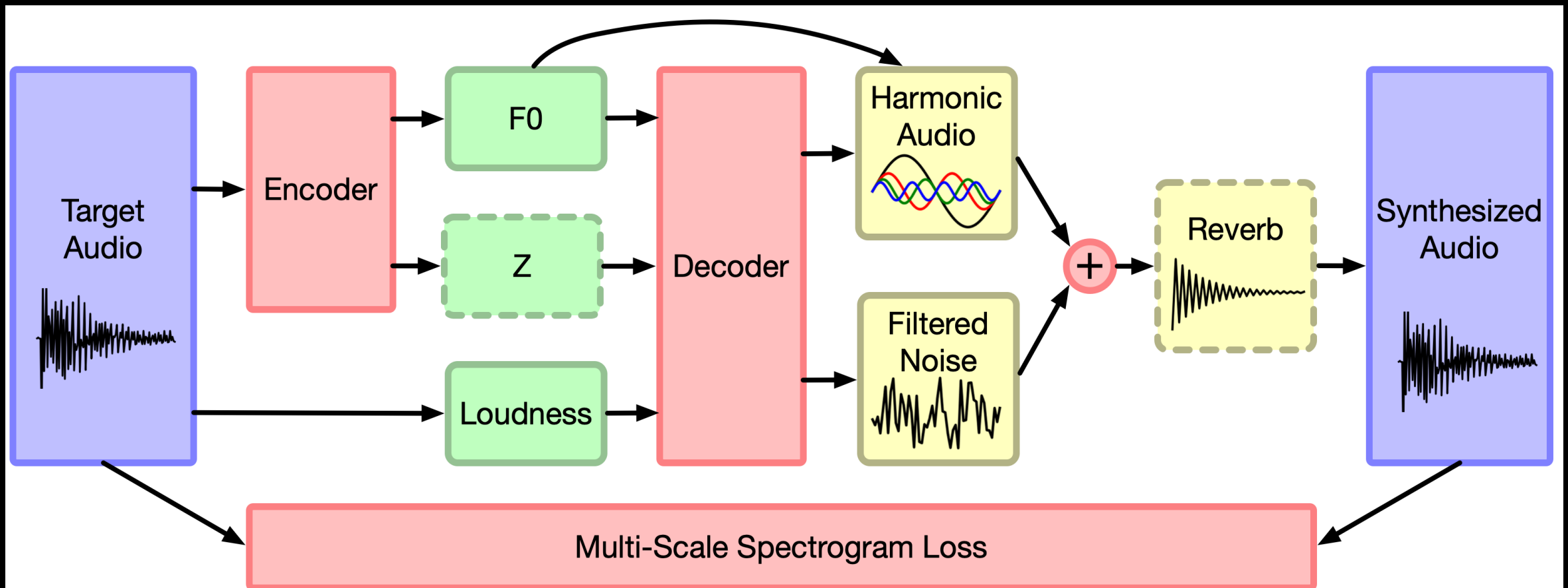
Additive
Synth



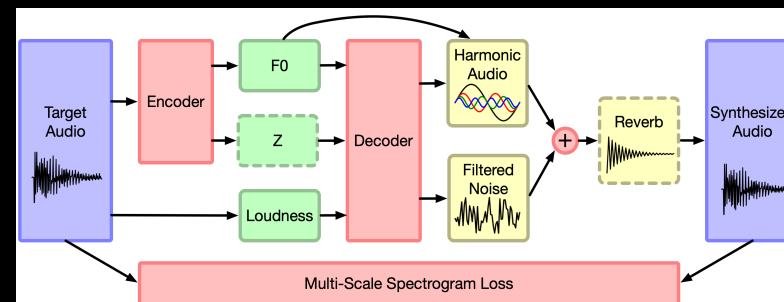
Filtered Noise
Synth



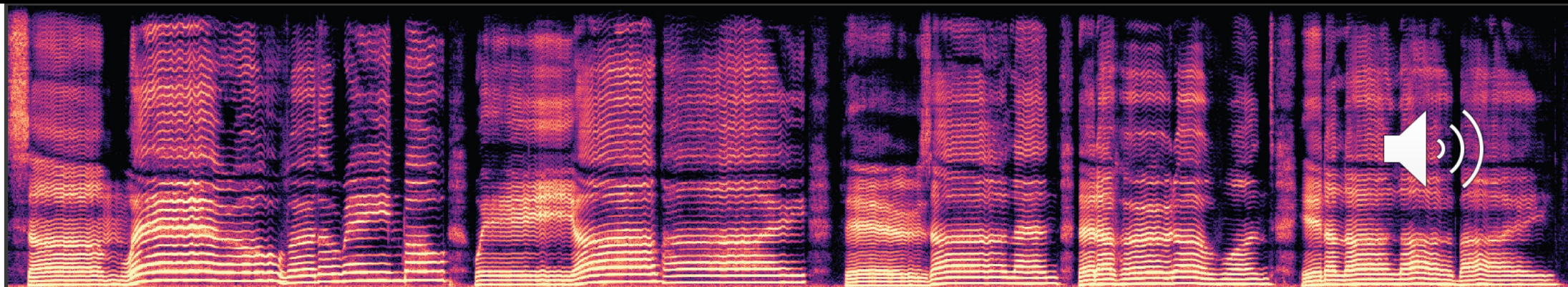
Music Audio Synthesis. DDSP



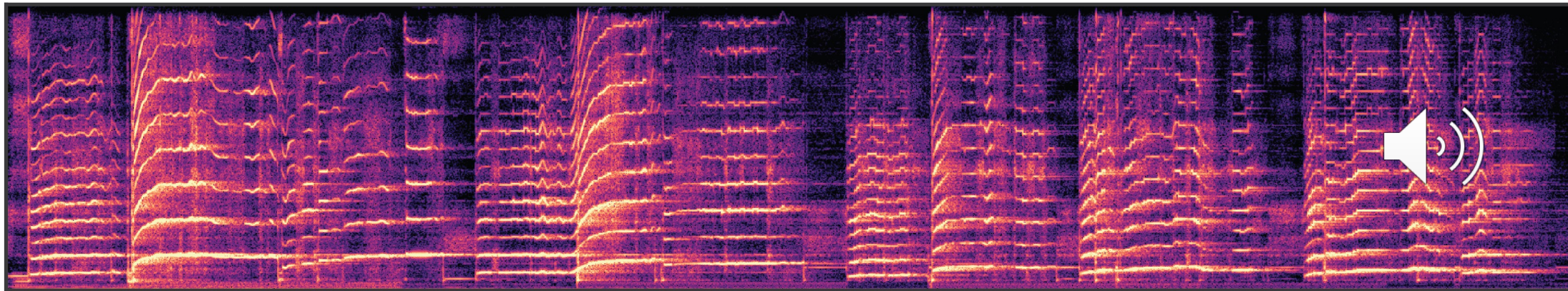
Music Audio Synthesis. DDSP



Original
(Singing)



Timbre Transfer
(Violin)



От звука к музыке

Умеем

1. Кодировать дискретный аудиосигнал в набор параметров
2. Синтезировать аудио из них

От звука к музыке

Умеем

1. Кодировать дискретный аудиосигнал в набор параметров
2. Синтезировать аудио из них

Хотим

1. Создавать «программу»
2. Превращать ее в музыку

От звука к музыке

Умеем

1. Кодировать дискретный аудиосигнал в набор параметров
2. Синтезировать аудио из них

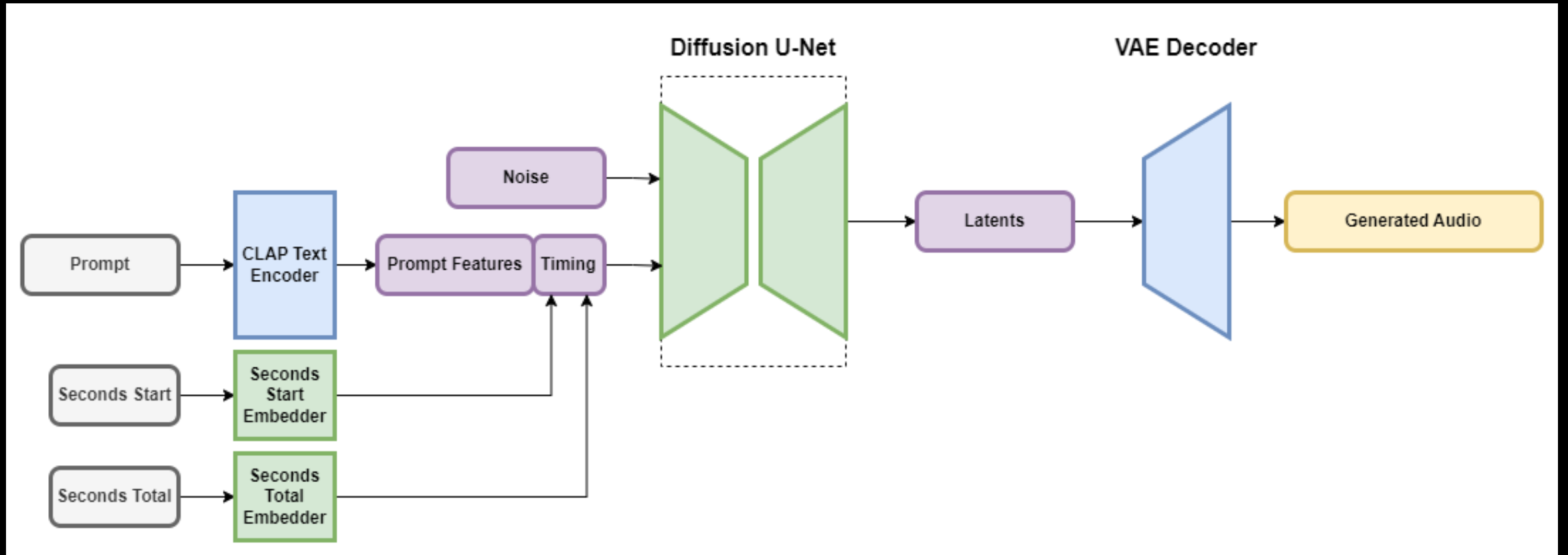
Хотим

1. Создавать «программу»
2. Превращать ее в музыку

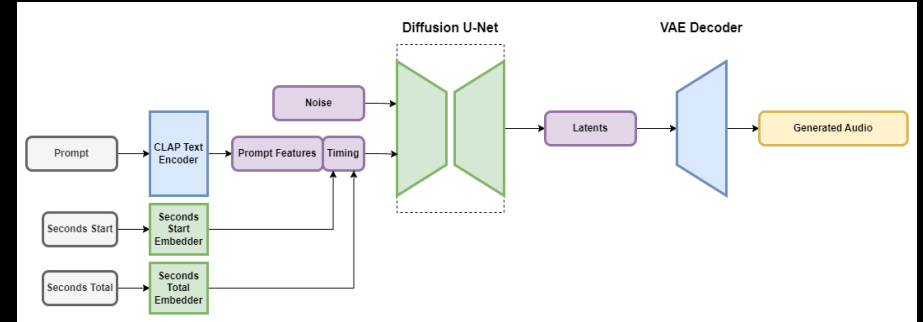
На основе

1. Имеющейся мелодии
2. Текстового запроса

Music Audio Generation. Stable Audio, 2024



Music Audio Generation. Stable Audio, 2024



Berlin techno, rave, drum machine, kick, ARP synthesizer, dark, moody, hypnotic, evolving, 135 BPM. Loop.

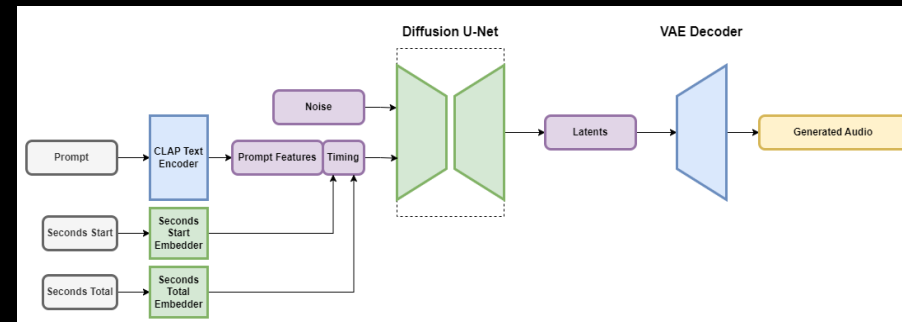


Calm meditation music to play in a spa lobby.

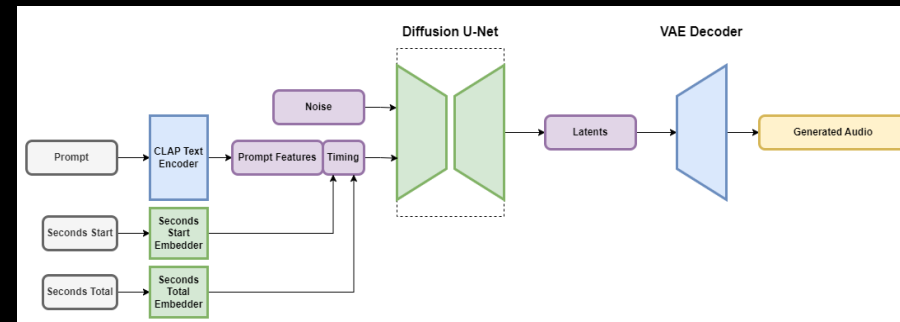


Music Audio Generation. Stable Audio, 2024

- Инструменты и звуки сразу в модели
- Не нужен пост-процессинг
- E2E
- Может давать музыку любых стилей



Music Audio Generation. Stable Audio, 2024



- Инструменты и звуки сразу в модели
- Не нужен пост-процессинг
- E2E
- Может давать музыку любых стилей



- Артефакты
- Качество звука
- Сложно редактировать результат
- Долго/дорого учатся и генерируют:
95с 44кГц за 8с на A100GPU
или 8минут на вашем M1pro

Music Generation



Идея!



Идея!

разделим синтез аудио
и сочинение нот!

Музыка механизмов. Музыкальная шкатулка

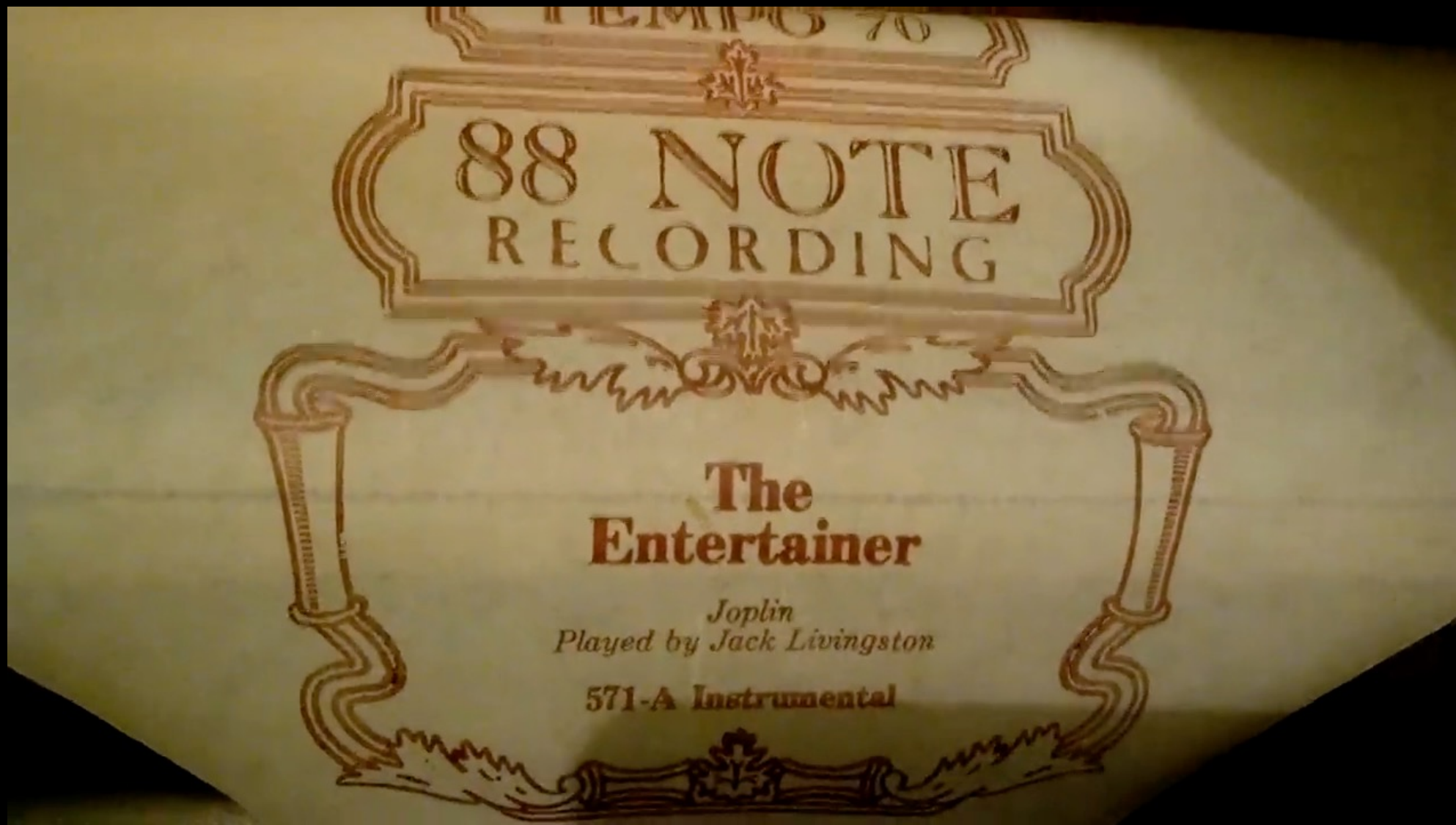
Антуан
Фавр,
1796



Музыка механизмов. Пианоролл, Оркестрион и Пианола

Эмиль
Велте,
1883

Эдвин
Скотт
Вотей,
1896



Пیانоролл в цифровых аудиостанциях



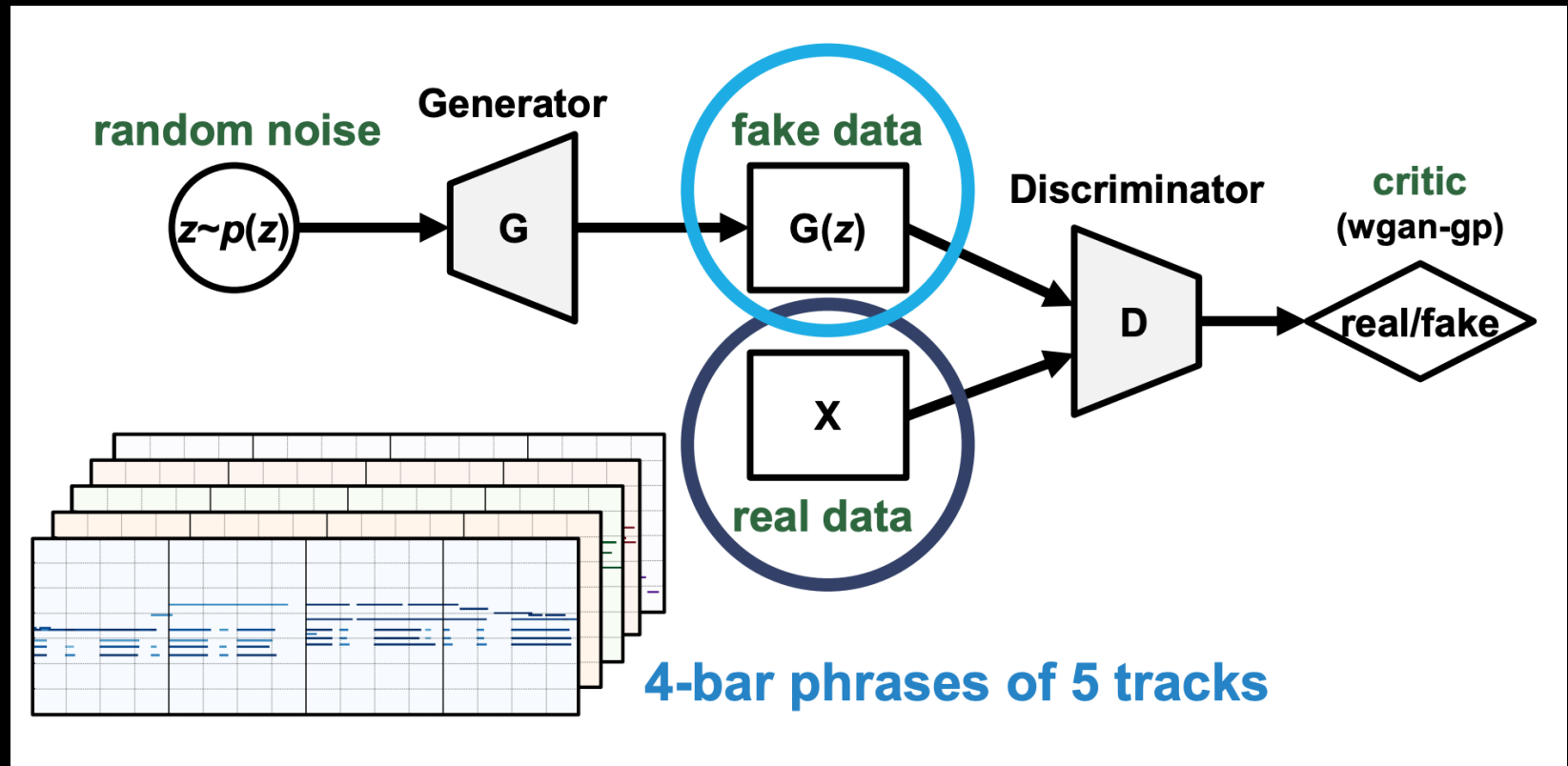


Идея!

Ноты как картинка!

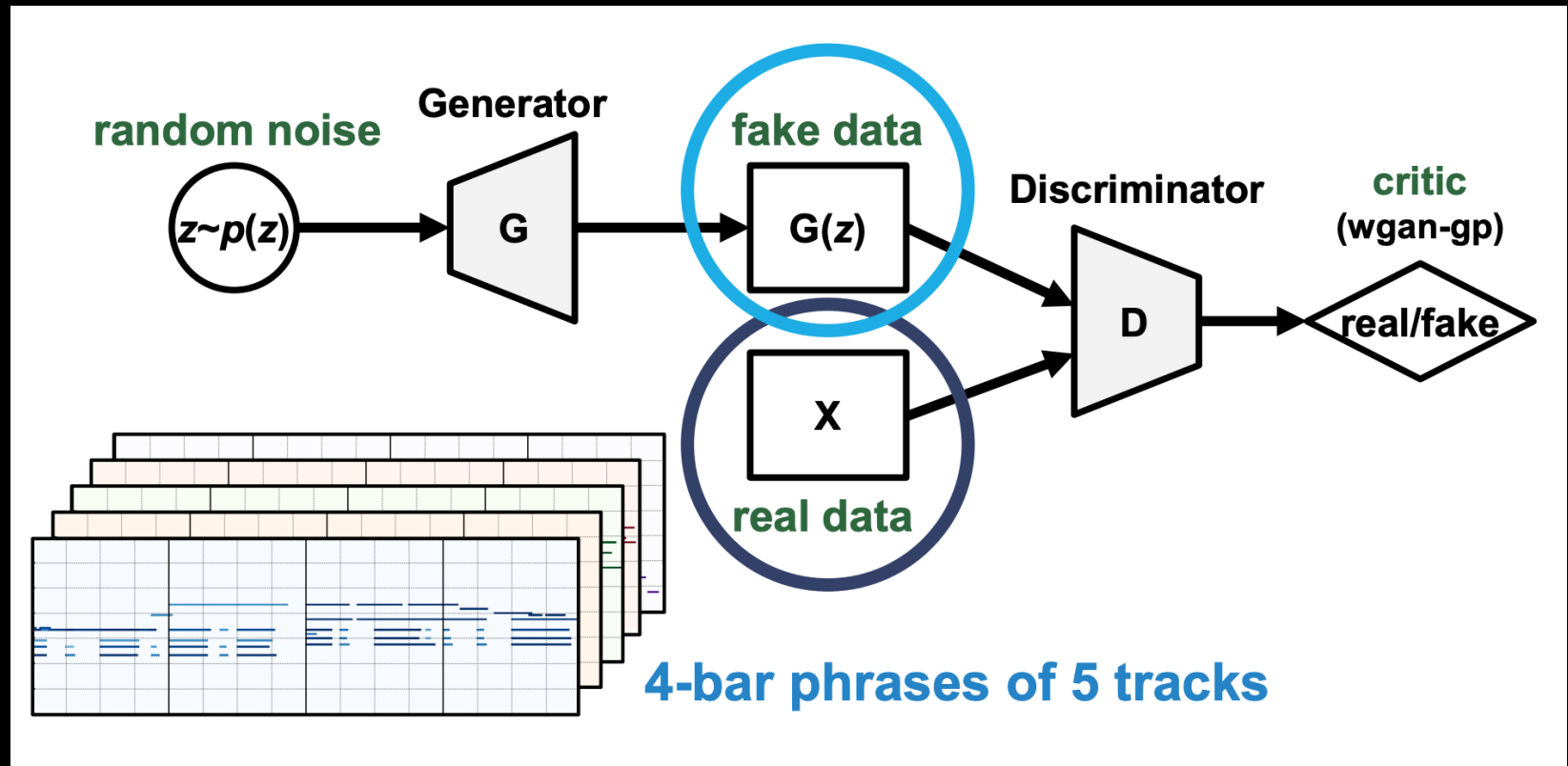
Pianoroll Generation. MuseGan, 2018

- Басс, Ударные, Гитара, Пианино и Струнные
- Только Поп/Рок
- 4 такта × 96 позиций × 84 высоты × 5 дорожек

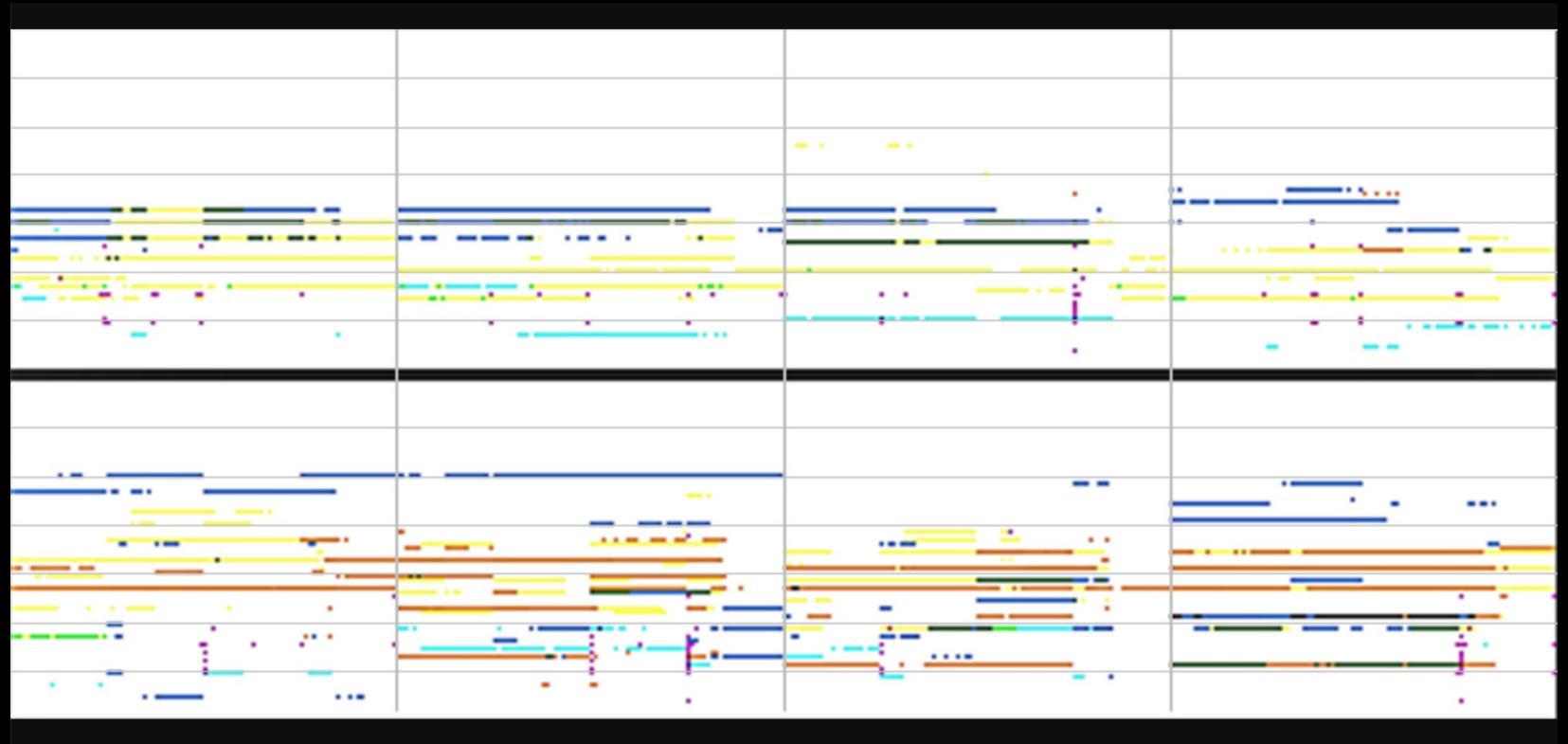


Pianoroll Generation. MuseGan, 2018

- Басс, Ударные, Гитара, Пианино и Струнные
- Только Поп/Рок
- 4 такта × 96 позиций × 84 высоты × 5 дорожек
- Сверточная нейросеть
- Генеративно-сопоставительное обучение



Pianoroll Generation. MuseGan, 2018

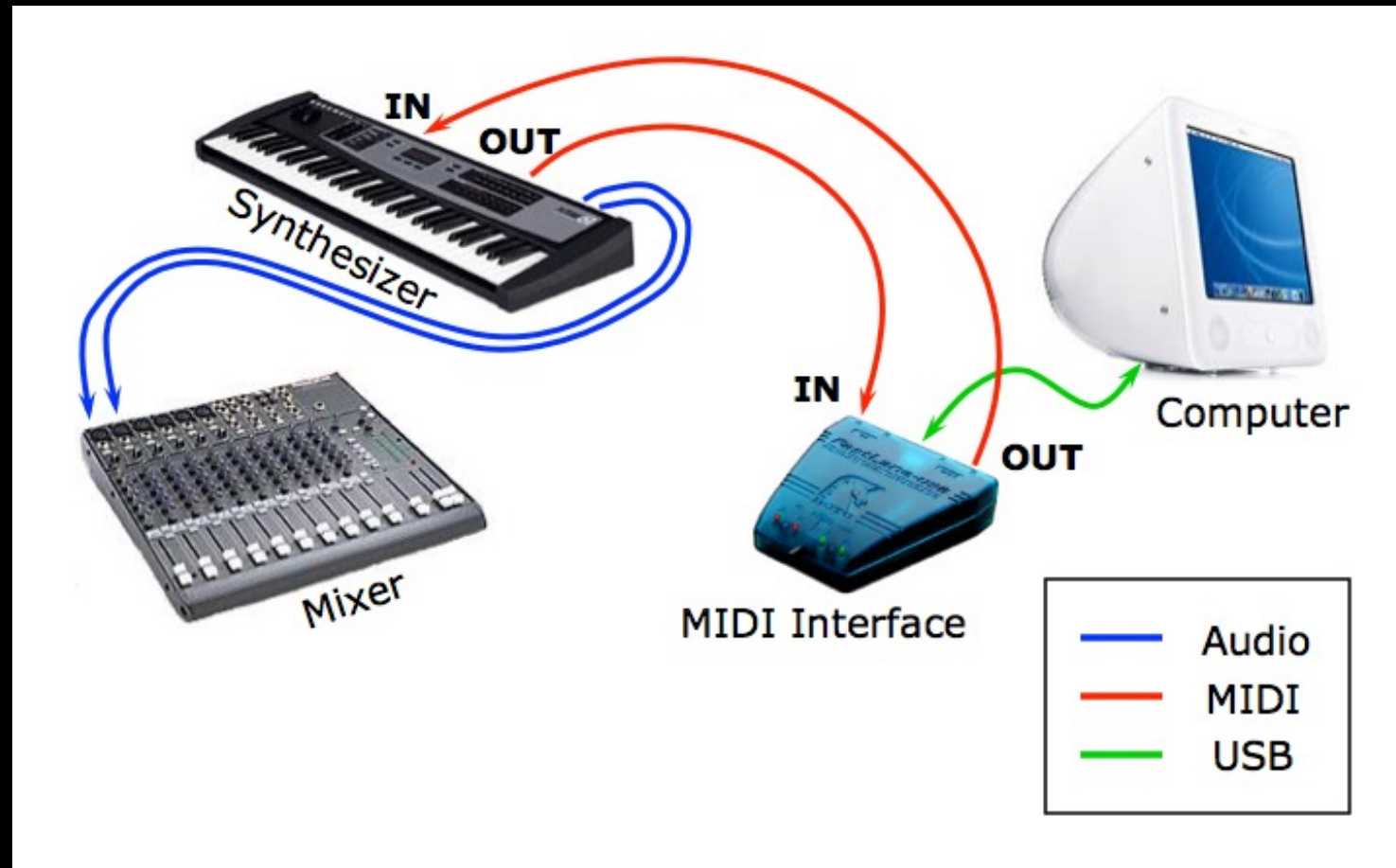


<https://salu133445.github.io/musegan/>

Hao-Wen Dong, Wen-Yi Hsiao, Li-Chia Yang, and YiHsuan Yang. MuseGAN: Multi-track Sequential Generative Adversarial Networks for Symbolic Music Generation and Accompaniment. In AAAI Conference on Artificial Intelligence, pages 34–41, 2018.

Символьный домен музыки. Ноты, Пианороллы, MIDI, MusicXml

MIDI,
1983



Символьный домен музыки. Ноты, Пианороллы, MIDI, MusicXml

НОТЫ <> MIDI

1 - Electric Piano 2



2 - Harp



3 - Bassoon



Delta-Time (decimal)	Event-Code (hex)	Other Bytes (decimal)	Comment
0	FF 58	04 04 02 24 08	4 bytes; 4/4 time; 24 MIDI clocks/click, 8 32nd notes/ 24 MIDI clocks (24 MIDI clocks = 1 crotchet = 1 beat)
0	FF 51	03 500000	3 bytes: 500,000 usec/ quarter note = 120 beats/minute
0	C0	5	Ch.1 Program Change 5 = GM Patch 6 = Electric Piano 2
0	C1	46	Ch.2 Program Change 46 = GM Patch 47 = Harp
0	C2	70	Ch.3 Program Change 70 = GM Patch 71 = Bassoon
0	92	48 96	Ch.3 Note On C3, forte
0	92	60 96	Ch.3 Note On C4, forte
96	91	67 64	Ch.2 Note On G4, mezzo-forte
96	90	76 32	Ch.1 Note On E5, piano
192	82	48 64	Ch.3 Note Off C3, standard
0	82	60 64	Ch.3 Note Off C4, standard
0	81	67 64	Ch.2 Note Off G4, standard
0	80	76 64	Ch.1 Note Off E5, standard
0	FF 2F	00	Track End



Идея!

Музыка как текст!



Идея!

Музыка как текст!



Symbolic Music Generation. REMI, 2020



Bar, **Position (1/16)**, **Chord (C major)**,
Position (1/16), **Tempo Class (mid)**,
Tempo Value (10), **Position (1/16)**,
Note Velocity (16), **Note On (60)**,
Note Duration (4), **Position (5/16)**,

Tempo Value (12), **Position (9/16)**,
Note Velocity (14), **Note On (67)**,
Note Duration (8), **Bar**

	MIDI-like	REMI (this paper)
Note onset	NOTE-ON (0-127)	NOTE-ON (0-127)
Note offset	NOTE-OFF (0-127)	NOTE DURATION (32th note multiples; 1-64)
Time grid	TIME-SHIFT (10-1000ms)	POSITION (16 bins; 1-16) & BAR (1)
Tempo changes	x	TEMPO (30-209 BPM)
Chord	x	CHORD (60 types)

Transformers is All You Need!

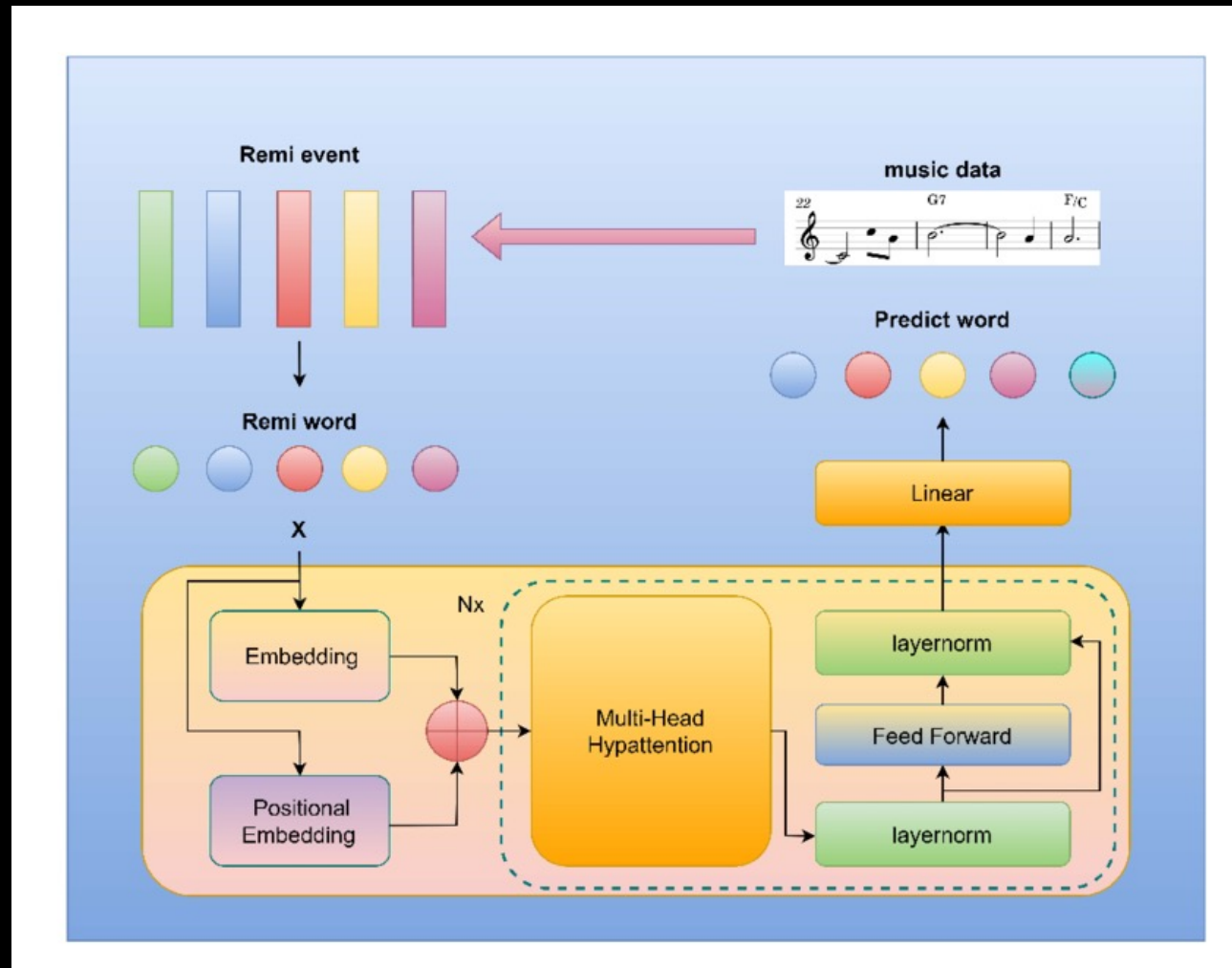


Bar, **Position (1/16)**, **Chord (C major)**,
Position (1/16), **Tempo Class (mid)**,
Tempo Value (10), **Position (1/16)**,
Note Velocity (16), **Note On (60)**,
Note Duration (4), **Position (5/16)**,

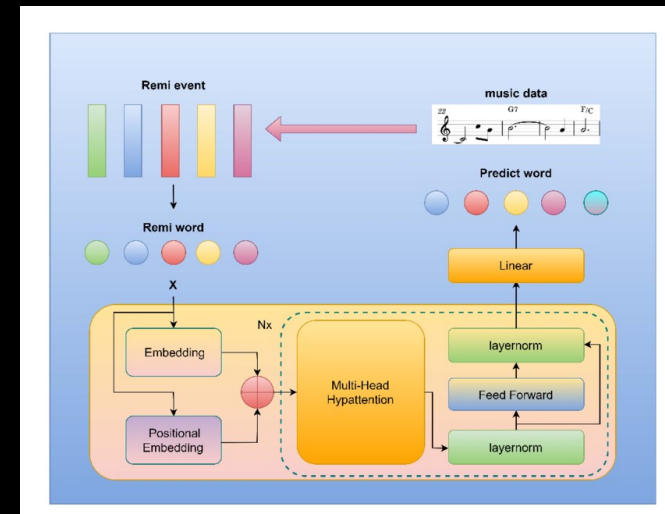
Tempo Value (12), **Position (9/16)**,
Note Velocity (14), **Note On (67)**,
Note Duration (8), **Bar**

	MIDI-like	REMI (this paper)
Note onset	NOTE-ON (0-127)	NOTE-ON (0-127)
Note offset	NOTE-OFF (0-127)	NOTE DURATION (32th note multiples; 1-64)
Time grid	TIME-SHIFT (10-1000ms)	POSITION (16 bins; 1-16) & BAR (1)
Tempo changes	x	TEMPO (30-209 BPM)
Chord	x	CHORD (60 types)

Symbolic Music Generation. REMI, 2020



Symbolic Music Generation. REMI, Transformers



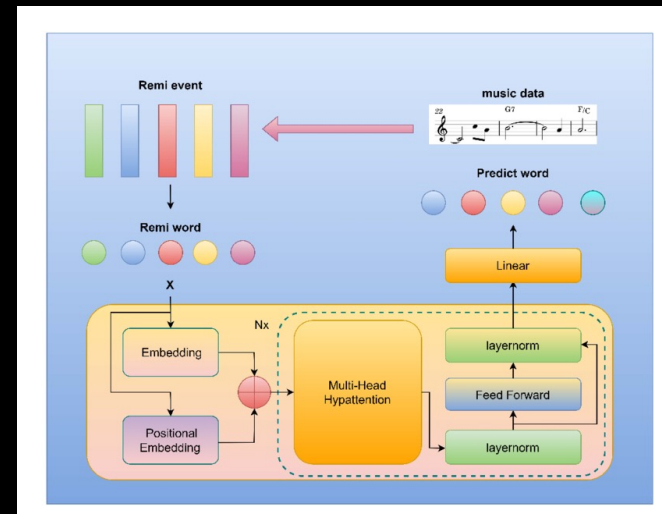
- Нужен дополнительный софт для рендера
- Звуковые эффекты и шумы – на стороне клиента
- Автоматические аранжировки могут сбиваться, если вход/датасет «грязный»

Symbolic Music Generation. REMI, Transformers

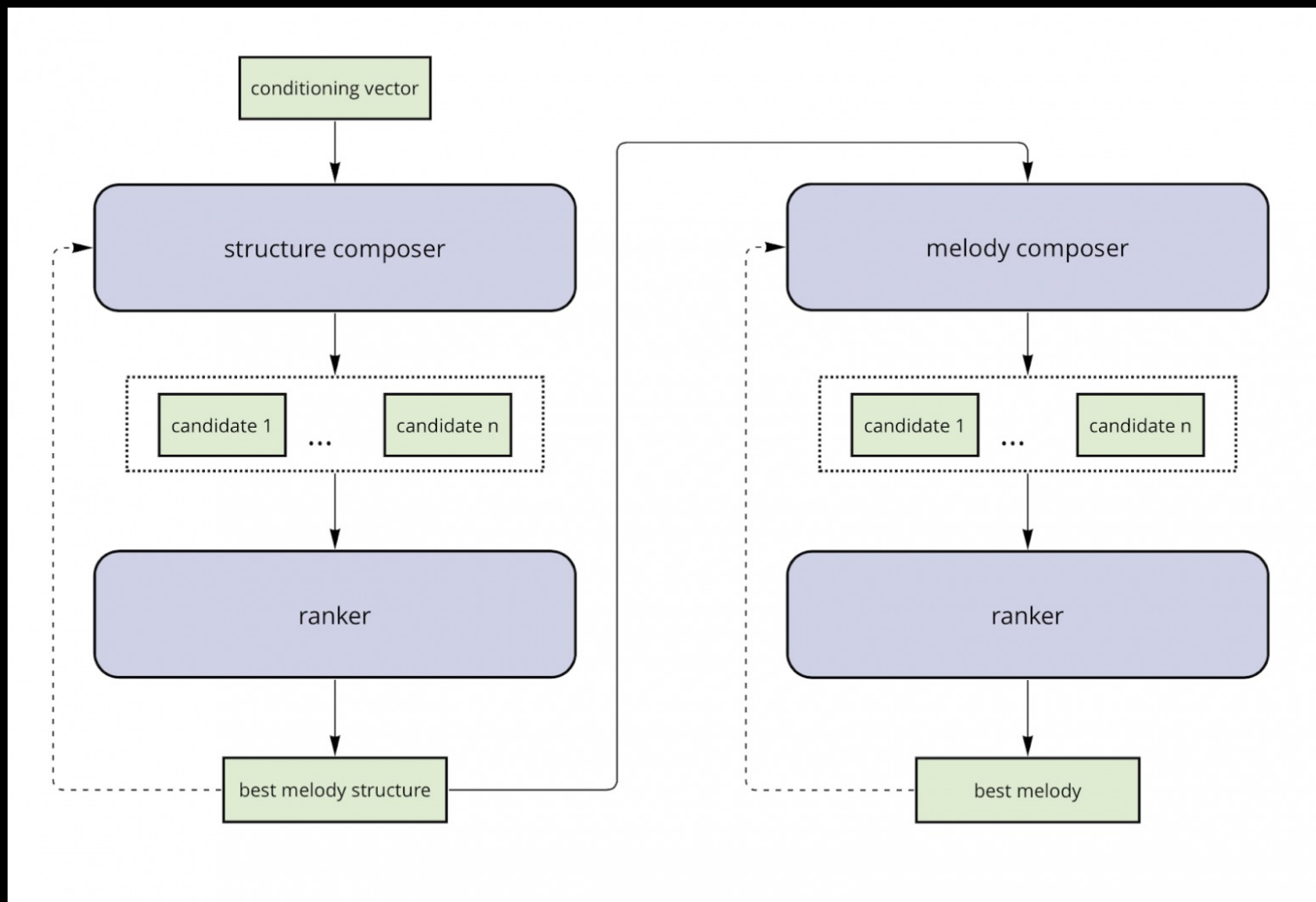
- Широкие возможности для контроля и модификаций
- Можно задавать инструменты
- Можно генерировать на основе задумки композиции
- Высокое качество звука с помощью VST



- Нужен дополнительный софт для рендера
- Звуковые эффекты и шумы – на стороне клиента
- Автоматические аранжировки могут сбиваться, если вход/датасет «грязный»

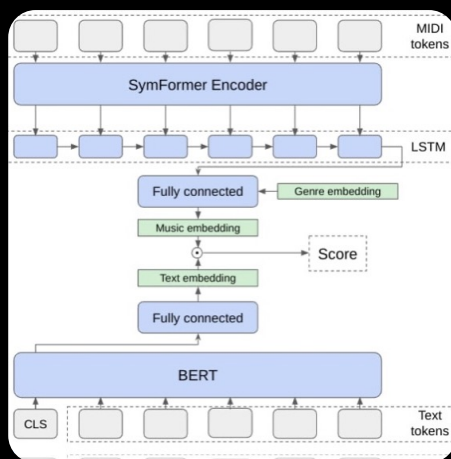


Symbolic Music Generation. SymFormer (SberDevices)

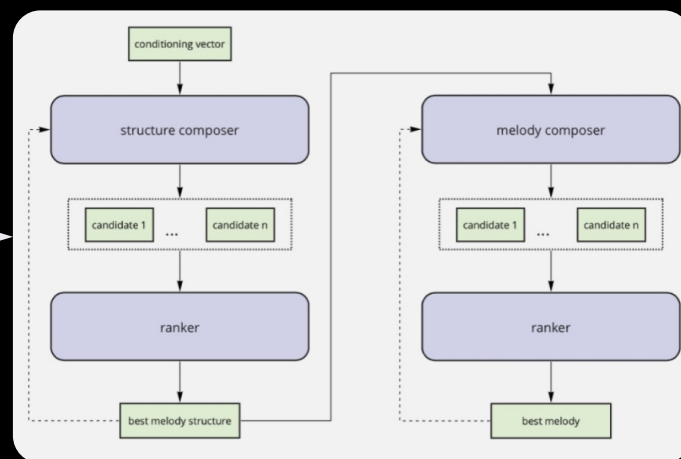


Сервис для генерации музыки. Maestro (SberDevices)

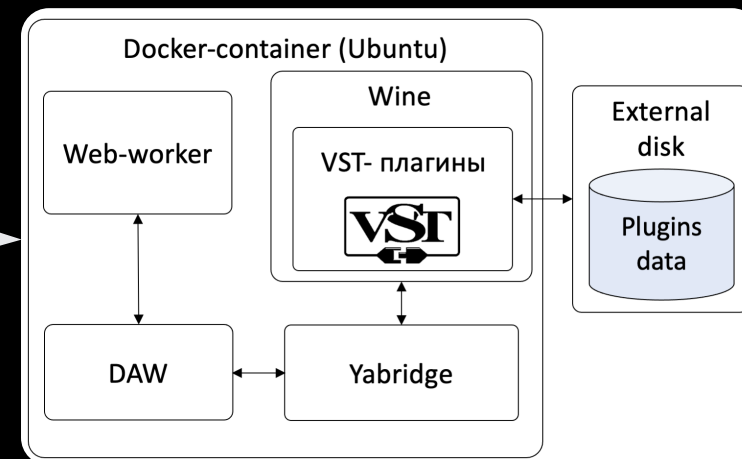
CLaMP



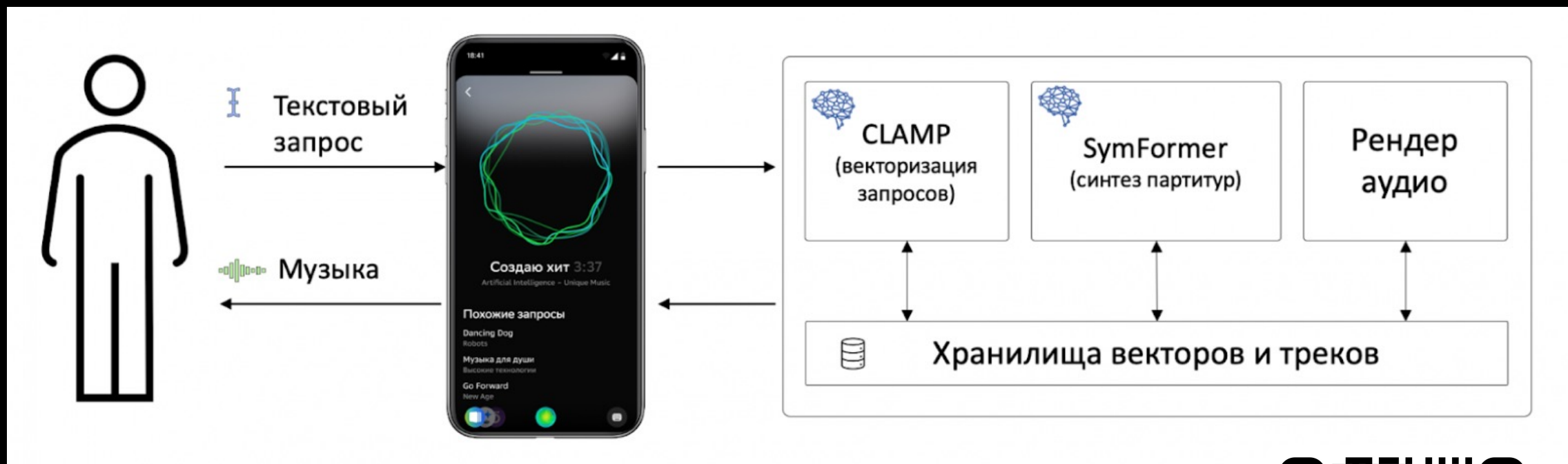
SymFormer



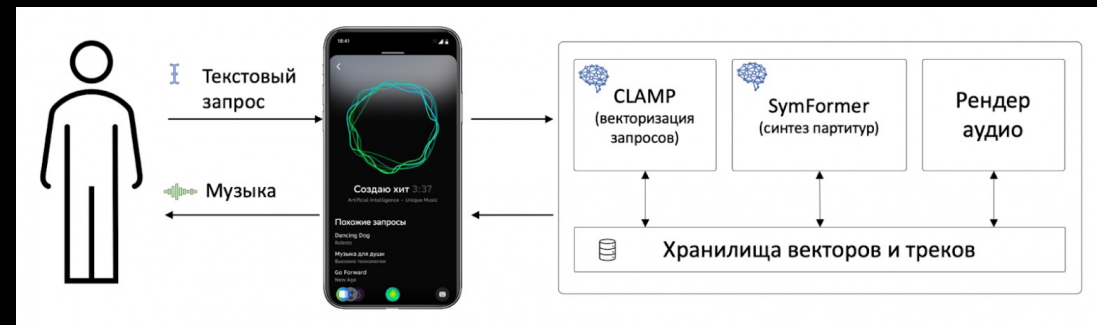
Renderer



Сервис для генерации музыки. Maestro (SberDevices)



Сервис для генерации музыки. Maestro (SberDevices)



Вольфганг
Амадей Моцарт –
Отрывок из
симфонии



Peace Makers –
Relax Take It
Easy



**ТЫ ВРОДЕ ВОКАЛОМ ПЛАНИРОВАЛ
ЗАНЯТЬСЯ? ЗАПИСАЛСЯ НАДЕЮСЬ?**



Нейросетевое пение

- Сгенерировать полноценный трек
- Расширить возможности музыкантов и композиторов
- Сделать общение с ассистентами еще более живым!



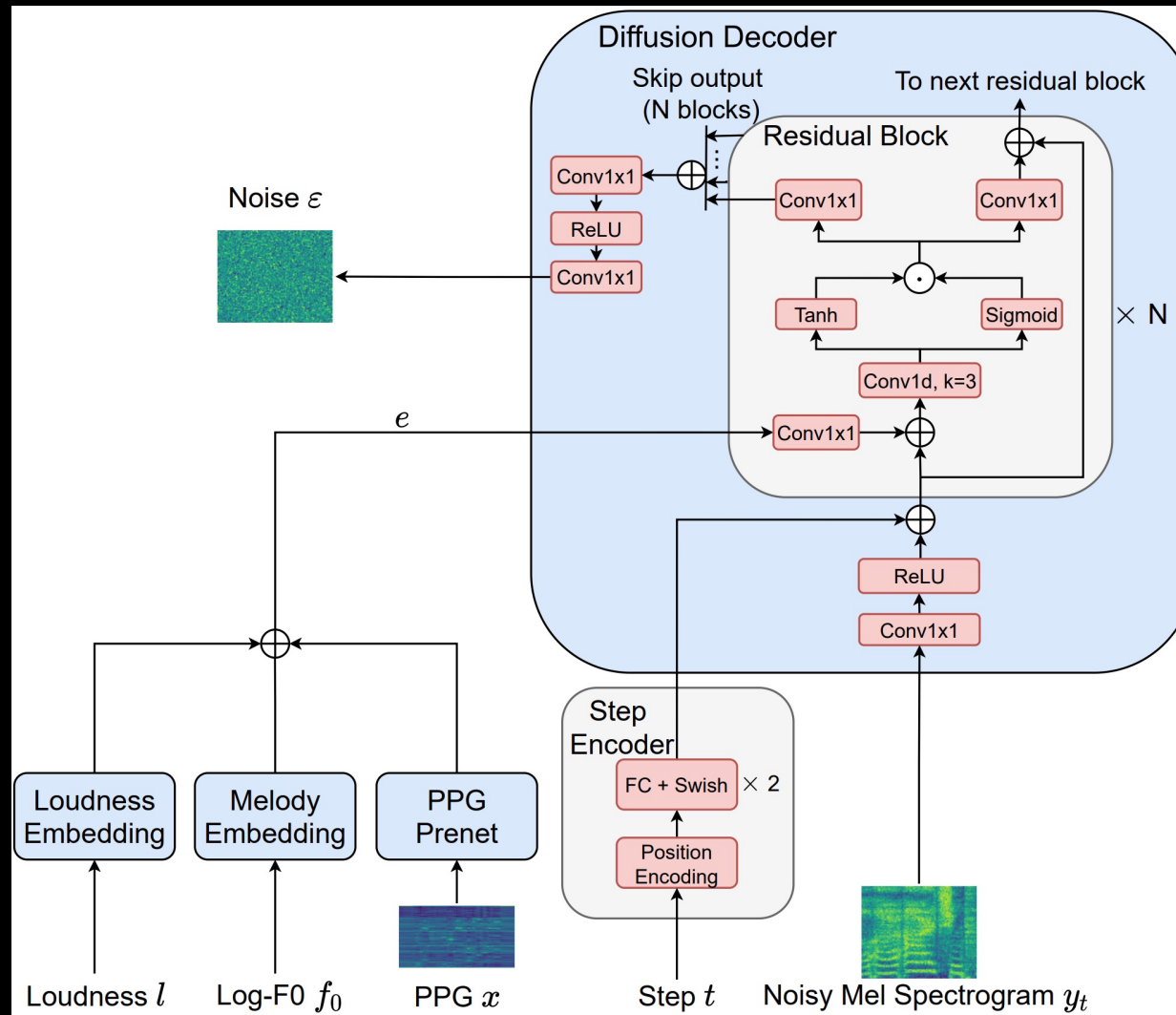
Нейросетевое пение

- Сгенерировать полноценный трек
- Расширить возможности музыкантов и композиторов
- Сделать общение с ассистентами еще более живым!



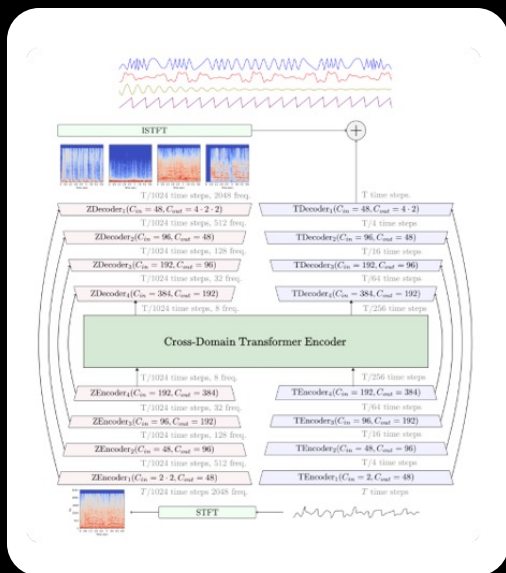
- Singing Voice Synthesis
- Singing Voice Beautifier
- Singing Voice Conversion

Singing Voice Conversion. DiffSVC



Singing Voice Conversion. AI Cover

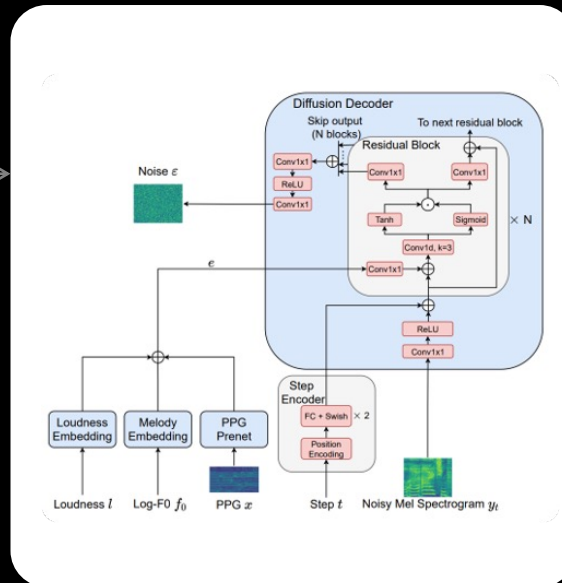
Original Track



Separator

Vocal

Instrumental



VC-model architecture

Vocal conversion

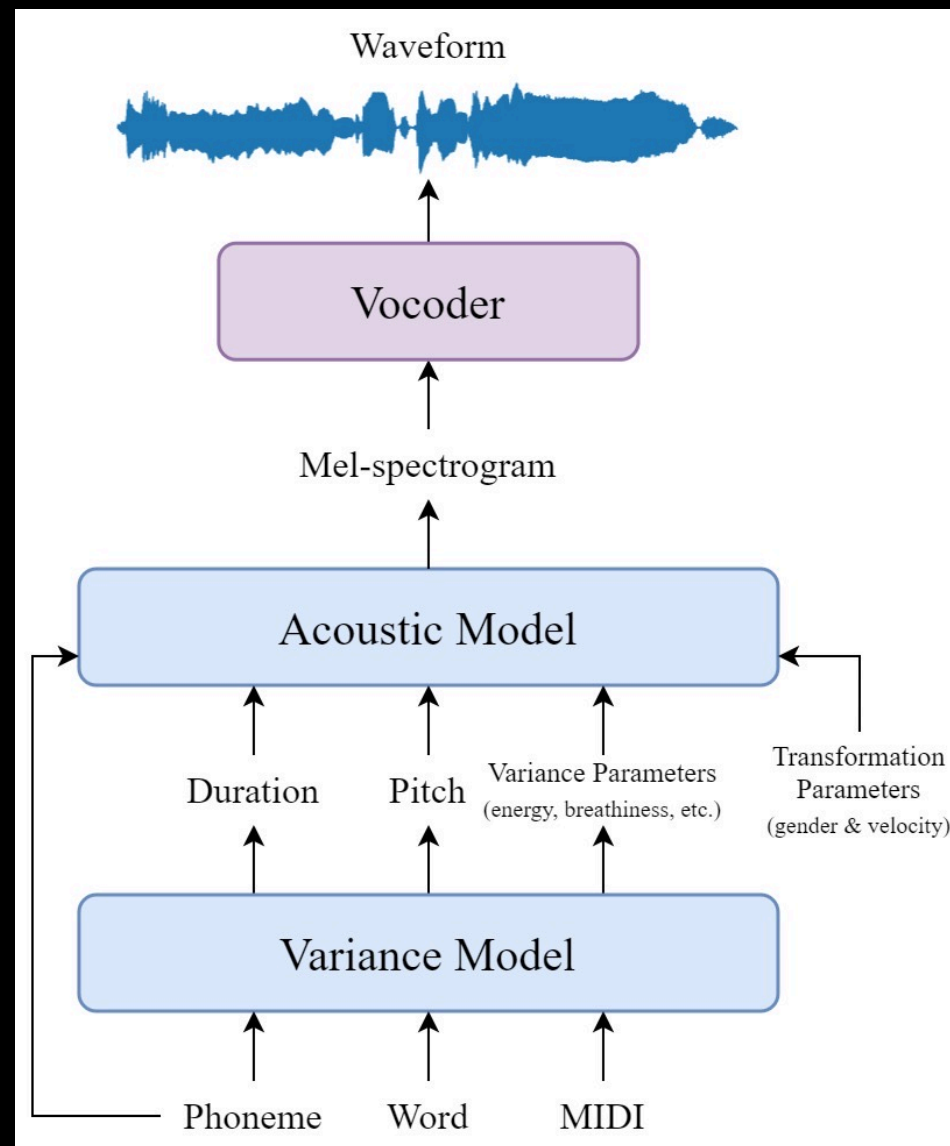
Speaker icon

Сбер, Афина, Джой

«Солнце Монако», Люся Чеботина

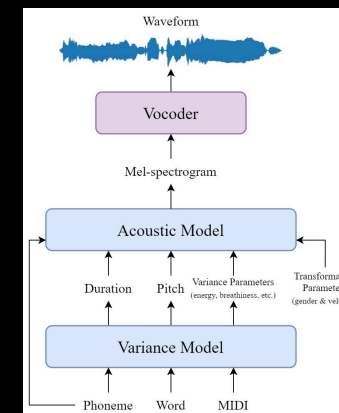
Singing Voice Synthesis

```
1  [
2      {
3          "offset": 0.0,
4          "text": "V LL I S UU R AY DD
5 I L AA SS J OO L AY CH K AY V LL I S UU A
6 N AA R AY S L AA",
7          "ph_seq": "V LL I S UU R AY
8 DD I L AA SS J OO L AY CH K AY V LL I S
9 UU A N AA R AY S L AA",
10         "ph_num": "3 2 2 2 3 2 3 2 3
11 2 1 2 3 2",
12         "note_slur": "0 0 0 0 0 0 0 0
13 0 0 0 0 0 0",
14         "note_seq": "B3 G#4 G#4 F#4
15 G#4 E4 B3 B3 B3 G#4 G#4 A4 C#5 B4",
16         "note_dur": "0.29741371875
17 0.4137930000000001 0.4051723125000001
18 0.40517231249999996 0.4568964375
19 0.3663792187499999 0.43534471875000014
20 0.39655162499999985 0.4224136875000002
21 0.40517231249999996 0.4094826562499998
22 0.41379300000000001 0.41810334375000036
23 0.8405170312500001"
24     }
25 ]
```



Singing Voice Synthesis

```
1  [
2      {
3          "offset": 0.0,
4          "text": "V LL I S UU R AY DD
5 I L AA SS J OO L AY CH K AY V LL I S UU A
6 N AA R AY S L AA",
7          "ph_seq": "V LL I S UU R AY
8 DD I L AA SS J OO L AY CH K AY V LL I S
9 UU A N AA R AY S L AA",
10         "ph_num": "3 2 2 2 3 2 3 2 3
11 2 1 2 3 2",
12         "note_slur": "0 0 0 0 0 0 0 0
13 0 0 0 0 0 0",
14         "note_seq": "B3 G#4 G#4 F#4
15 G#4 E4 B3 B3 B3 G#4 G#4 A4 C#5 B4",
16         "note_dur": "0.29741371875
17 0.4137930000000001 0.4051723125000001
18 0.40517231249999996 0.4568964375
19 0.3663792187499999 0.43534471875000014
20 0.39655162499999985 0.4224136875000002
21 0.40517231249999996 0.4094826562499998
22 0.41379300000000001 0.41810334375000036
23 0.8405170312500001"
24     }
25 ]
```



Оригинальный текст

В лесу родилась ёлочка,
В лесу она росла.
Зимой и летом стройная,
Зеленая была.



Собираем вместе!

«Взгляд в будущее»
Композитор SymFormer
Поющая нейросеть от
Салют
Замена голоса:
Джой, Сбер, Афина



Сгенерированный текст

- Я путешествую в мирах виртуальных
- Я разбираюсь в интерфейсах нейральных
- А сейчас пою вам эту песню – нейропривет!

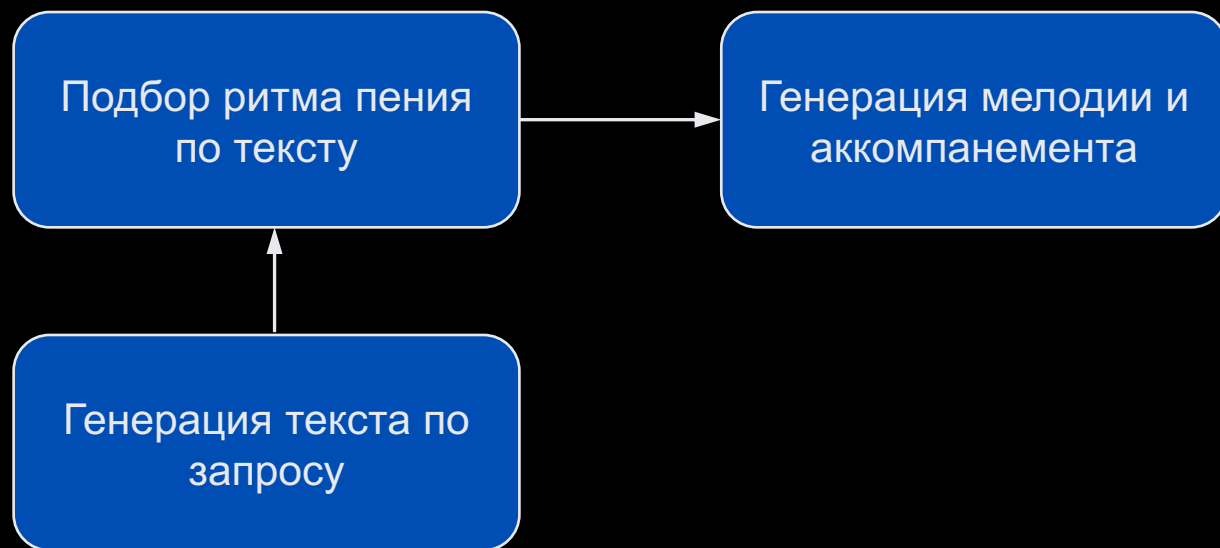
Автоматизированное создание музыкального трека!

Генерация текста по
запросу

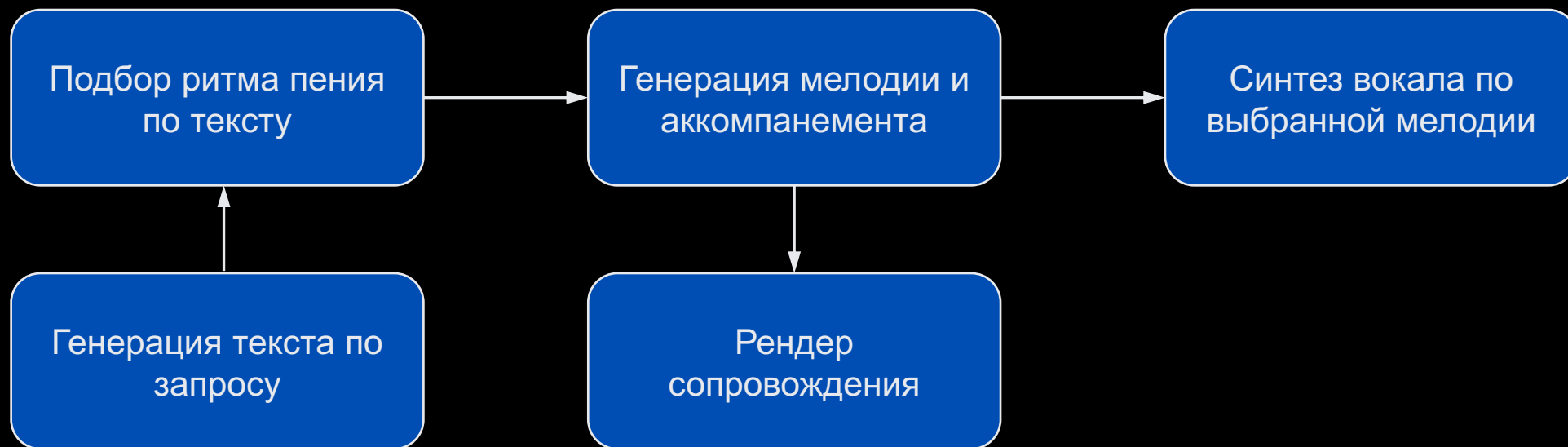
Автоматизированное создание музыкального трека!



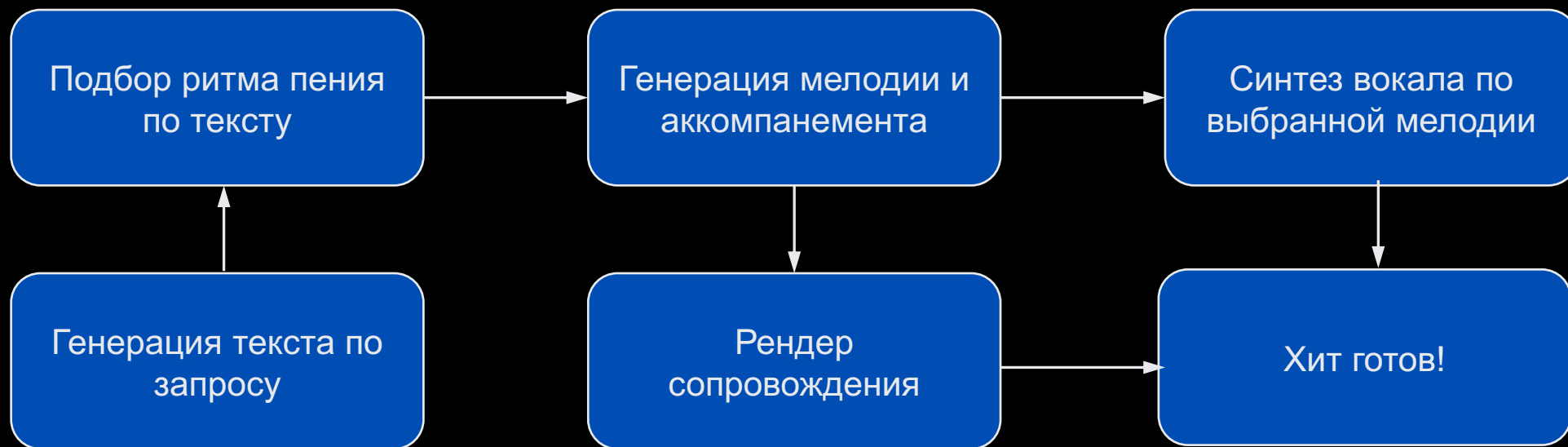
Автоматизированное создание музыкального трека!



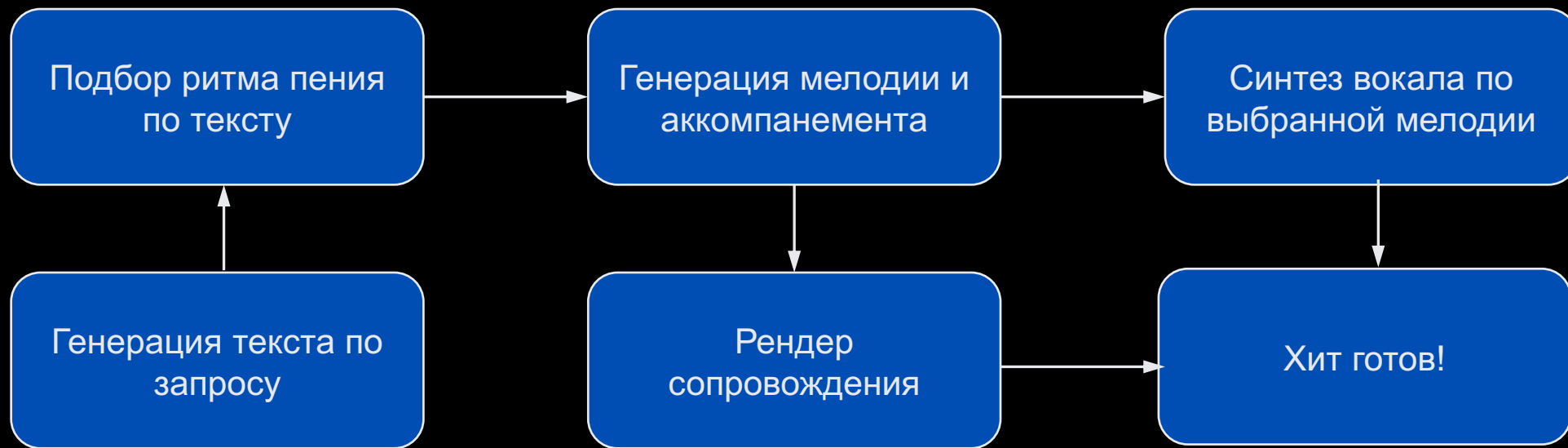
Автоматизированное создание музыкального трека!



Автоматизированное создание музыкального трека!



Автоматизированное создание музыкального трека!



GigaChat

+ Lyrics2Melody + SymFormer

+ SVS + Render

Автоматизированное создание музыкального трека!

В небе парит буревестник,
Гарри Поттер в небесах.
Все что от него осталось,
Это память в наших сердцах.

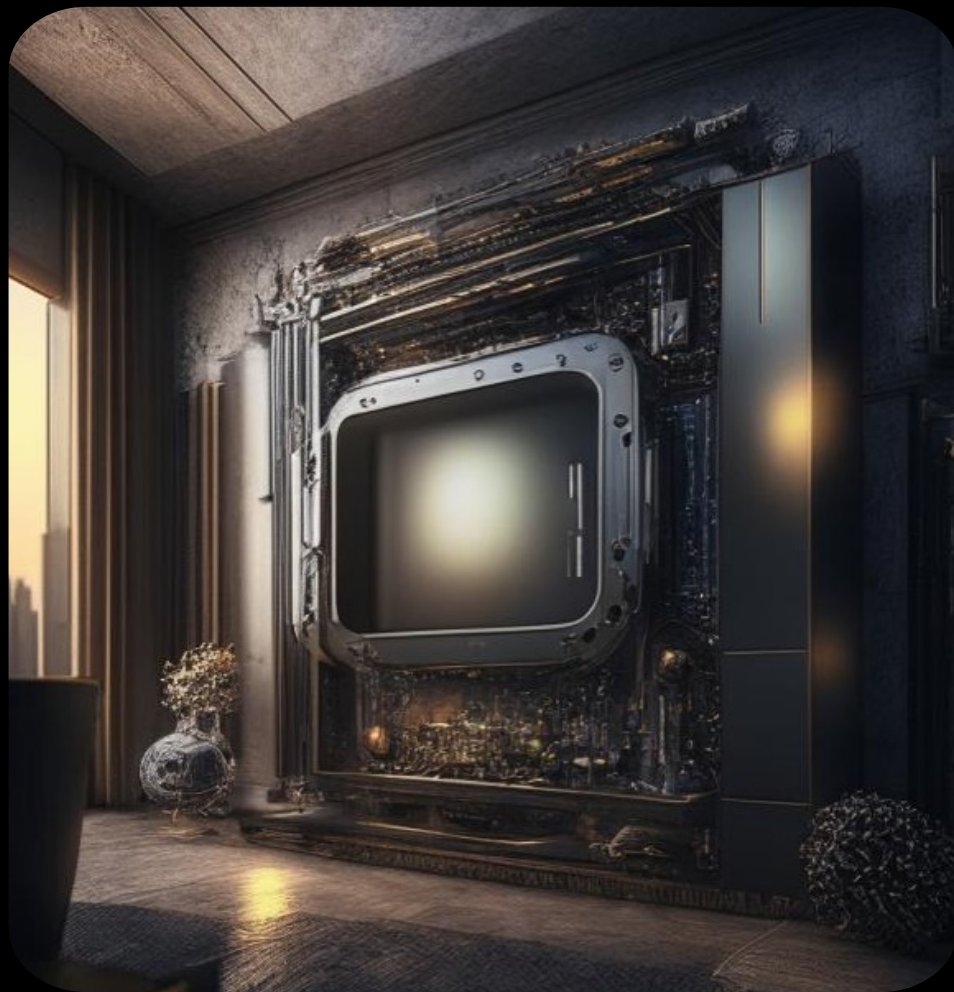


GigaChat

+ Lyrics2Melody + SymFormer

+ SVS + Render

Больше интересных приложений!

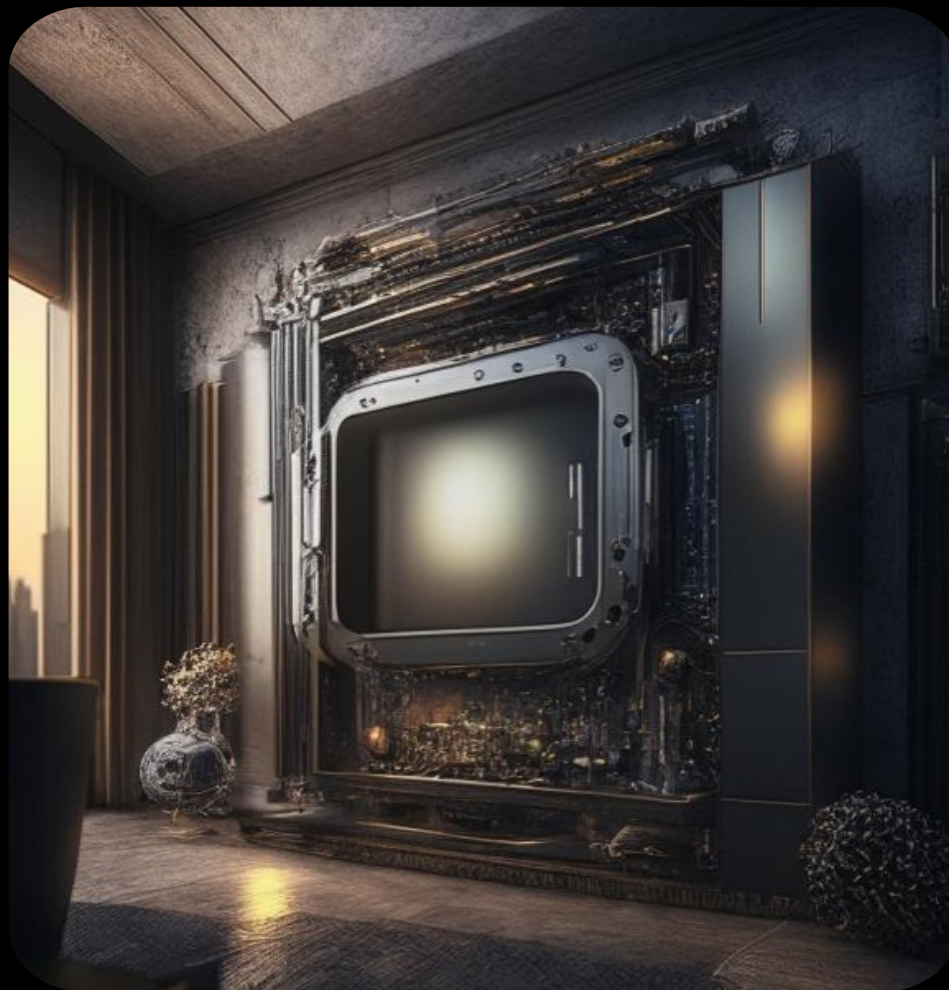


Больше интересных приложений!

Выставка
Роботов

AI Journey
2023

Московский
урбанистический
форум 2023



Открытия
«Школы21»

Встречи
почетных
гостей

Внутренние
мероприятия

Симфония №1 «Космос»

для искусственного интеллекта с Большим симфоническим оркестром

Композитор и дирижёр:
Пётр Дранга

Импровизация:
SymFormer от Сбера

Впервые представлена:
10.09.2023 в рамках
культурной программы ВЭФ



Svoboda Zvuka & SymFormer

Сейшен-импровизация с
музыкальным
коллективом

Real-time streaming из
нейросетевого композитора
SymFormer

Giga RnD Day, 05.04.2024



Наша музыка в ЗВУК



Aphoria



12 релизов





Бонус! Подборка приложений и моделей с генеративной музыкой!



Maestro



<https://gitverse.ru/mintas/awesome-music-ai>
<https://github.com/Mintas/awesome-music-ai>



Aiphoria