VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS
PROGRAMŲ SISTEMŲ BAKALAURO STUDIJŲ PROGRAMA

# Muzikos generavimas naudojant dirbtinį intelektą ir Blockly sąsają

## Music Generation Using Artificial Intelligence and Blockly Interface

Bakalauro baigiamasis darbas

Atliko: Aurimas Adlys

Darbo vadovas: lekt. Irus Grinis

Darbo recenzentas: asist. dr. Vytautas Valaitis

Vilnius – 2024

# Santrauka

Šiame darbe buvo tiriamas muzikos generavimas naudojant dirbtinį intelektą. Išanalizuotos trijų tipų modelių (LSTM, VQ-VAE, transformerių) galimybės atliekant šią užduotį. Taip pat pasiūlytas būdas objektyviai kokybiškai vertinti tokią muziką. Atlikti du eksperimentai, su minėtais modeliais generuojant dviejų žanrų - klasikinės bei pop muzikos - kūrinius bei juos įvertinant, tuo pačiu lyginant su žinomais šių žanrų kūriniais. Galiausiai, pasiūlytas dirbtinio intelekto generuojamos muzikos integracijos į Blockly pagrindu veikiančią programą konceptas.

**Raktiniai žodžiai: Dirbtinis intelektas, muzikos generavimas, muzikos kokybės vertinimas, Blockly**

# Summary

This paper studies music generation using artificial intelligence. Three types of models' (LSTM, VQ-VAE, transformers) capabilities on this task were analyzed. The paper also suggests a method to objectively evaluate the quality of music generated by artificial intelligence. Two experiments were carried out, where the before-mentioned models were used to generate music of two genres - classical and pop and the outputs were evaluated while comparing the resuts to known musical pieces of each genre. Finally, the paper suggests a concept of the integration of music generated by artificial intelligence into an application built on Blockly.



**Keywords: Artificial intelligence, music generation, music quality evaluation, Blockly**

# Turinys

# Įvadas

## Tyrimo objektas ir aktualumas

Šio tyrimo pagrindinis objektas - muzikos generavimas naudojant dirbtinį intelektą. Dirbtinis intelektas - sritis, itin stipriai pažengusi pastaraisiais metais. Ši pažanga sukūrė galimybių įvairiose srityse, kaip edukacija [ZA21] bei medicina [LGL+21]. Tačiau, atsižvelgiant į pastovų bei spartų dirbtinio intelekto srities augimą bei tobulėjimą, autoriaus nuomone yra svarbu ištirti jo įtaką ne tik mokslui bei verslui, bet taip pat ir menui. Tą grindžia ir faktas, jog auga pasaulinio masto susidomėjimas būtent dirbtiniu intelektu grįstu muzikos generavimu. Šia tema publikacijos skelbiamos tiek akademinių, tiek privačių sektorių, prie to prisidedant ir dirbtinio intelekto sferos lyderiams, tokiems kaip „Google“, „OpenAI“ bei „Amazon“ [CCC+22].

Gilinimasis į dirbtinio intelekto generuojamą muziką bei jos tobulinimas autoriaus manymu - itin svarbus žingsnis šiuolaikiniame muzikos kūrime. Pagrindinė to priežastis - ta, jog dirbtinio intelekto modeliai mokydamiesi padengia įvairius nagrinėjamų stilių aspektus ir gali išmokti atpažinti bei išnaudoti ryšius tarp jų geriau, nei žmonės [KFV20].

Nagrinėjant šią temą, svarbu išanalizuoti ir sunkumus, su kuriais šiuo metu susiduriama bei įvardinti tobulėjimo galimybes. Vienas iš tokių iššūkių - apmokyti modelius, kurie gebėtų generuoti muziką su griežta struktūra, kuri ilgainiui neiširtų [HB22]. Kitos problematiškos sritys - kaip suteikti naudotojui pakankamai kontrolės, ir, kaip suteikti modeliui pakankamai kūrybiškumo [BP20]. Išvardintos problemos didžiąja dalimi reiškia, jog reikia būdų dirbtinio intelekto generuojamą muziką įvertinti kokybiškai. Tokius būdus iš esmės galima skaidyti į subjektyvius - besiremiančius žmogiška klausa bei vertinimu ir objektyvius, kai vertinamos įvairios metrikos - tono bei ritmo, harmonijos, stiliaus [XWY+23]. Darbe bus siekiama ir to - įvertinti dirbtinio intelekto generuojamos muzikos kokybę.

Tęsiant kursinio darbo projektą, bus bandoma pasiūlyti būda integruoti pasitelkiant dirbtinį intelektą generuojamą muziką į Blockly pagrindu pradėtą kurti įrankį muzikos kompozicijai. Šio įrankio paskirtis bei aktualumo sritis išlieka ta pati - jis kuriamas pagal edukacinio įrankio idėją.

### Darbo tikslas

Šio darbo tikslas - išanalizuoti bei palyginti skirtingus metodus generuoti muziką naudojant dirbtinį intelektą ir įvertinti tokią muziką kokybiškai bei rasti metodą integruoti šias technologijas į Blockly sąsają.

### Keliami uždaviniai

1. Ištirti dabartines dirbtinio intelekto galimybes generuojant muziką.
2. Išanalizuoti bei palyginti skirtingus metodus - modelius/algoritmus, naudojamus šiam tikslui.
3. Sukūrus arba naudojant sukurtą(-us) modelį(-ius), sugeneruoti muzikinių kūrinių bei juos kokybiškai įvertinti.

4. Rasti metodą integruoti dirbtinio intelekto generuojamą muziką į Blockly pagrindu veikiančios programą.

**Laukiami rezultatai**

1. Aptarta dirbtinio intelekto įtaka muzikos kūryboje.
2. Išanalizuoti bei palyginti skirtingi algoritmai/modeliai muzikos generavimui.
3. Pateiktas kokybinis dirbtinio intelekto generuojamos muzikos vertinimas.
4. Pasiūlytas metodas integruoti dirbtinio intelekto generuojamą muziką į Blockly pagrindu veikiančią programą.

**Metodai**

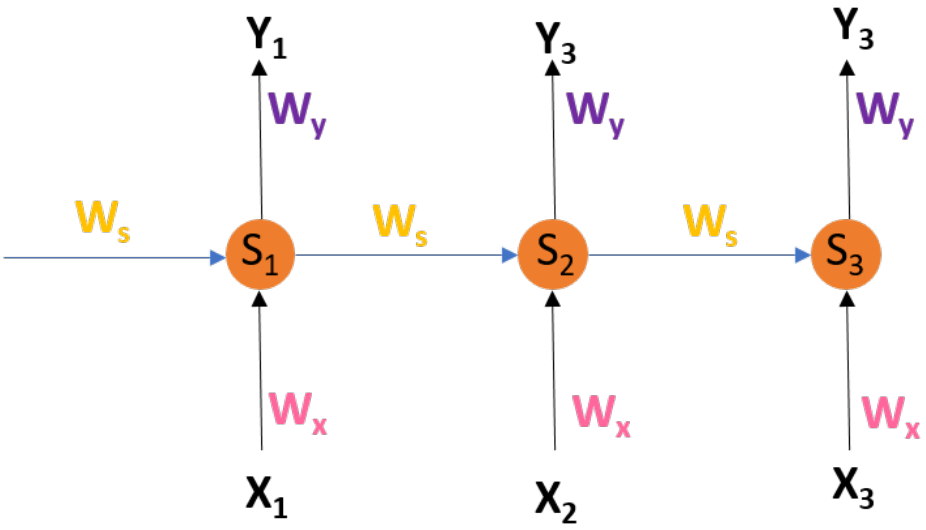
Šiam tyrimui naudojami trys žinomi dirbtinio intelekto modeliai (Jukebox, MuseNet, MusicRNN) muzikos generavimui, taip pat Python kalba sukurta programinė įranga muzikos vertinimo pagalbinei informacijai gauti, naudojama lyginamoji analizė.

# 1. LSTM

LSTM (angl. *Long Short-Term Memory*) modeliai yra rekurentinių neuroninių tinklų (angl. *recurrent neural network (RNN)*) (toliau darbe „RNN“) tipas. Todėl, kalbant apie LSTM, svarbu paminėti bendras rekurentinių neuroninių tinklų savybes.

## 1.1. RNN

RNN yra ypatingai naudingi darbui su duomenų sekomis, kadangi pasižymi veikimu, kurio metu saugomos paslėptos būsenos (angl. *hidden states*), padedančios rekursyviai apdoroti įvesčių sekas - išvestis priklauso ne tik nuo įvesties, tačiau ir nuo būsenos (atminties, kuria atsižvelgiama į praeitas įvestis). Dėl būtent šių požymių RNN yra tinkami muzikos generavimo užduotims, kadangi jas atliekant yra itin svarbu atsižvelgt į elementų laikinę priklausomybę nuo vienas kito, kuri daro didelę įtaką generuojamo rezultato (šiuo atveju - muzikos) nuoseklumui.



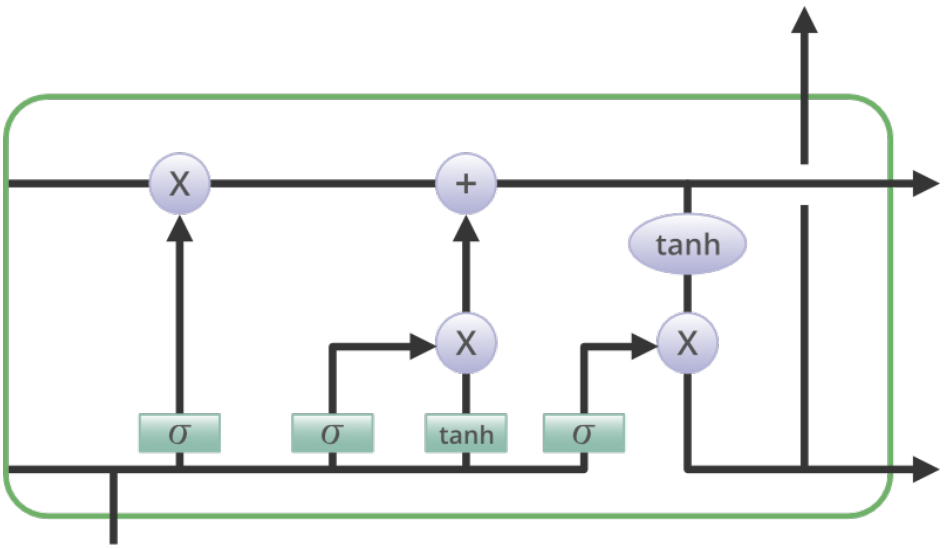

1 pav. RNN architektūros schema. Čia: Wx - įvestis, S - būsena, Ws - su ja susieta svorių matrica, Wy - išvestis tam tikru laiko momentu.

Tačiau įprasti RNN tinklai, ypač esant ilgam modelio mokymo procesui, susiduria su nykstančio gradiento problema (angl. *vanishing gradient problem*), kuri trikdo tinklo gebėjimą mokytis, ypač esant ilgalaikėms laiko priklausomybėms. Šis reiškinys susijęs su RNN naudojamu BPTT (angl. *Backpropagation Through Time*) algoritmu, kuris skirtas nuostolių funkcijos gradientui skaičiuoti [BSF94].

LSTM tinklų, kurie yra RNN tipas, vienas iš sukūrimo tikslų yra būtent šios problemos sprendimas.[HS97]. LSTM tinklai naudoja atminties ląsteles (angl. *memory cells*), galinčias išsaugoti informaciją per ilgą laiko tarpą.

## 1.2. LSTM architektūra

Dažnai LSTM tinklų architektūra turi tokią formą[SJ19]:




2 pav. LSTM architektūra

Įvestis $x_t$ laiko momentu $t$ apjungiama su buvusia paslėpta būsena $h_{t-1}$ į vieną vektorių, kuris paduodamas kaip parametras į formules: $f_t$ - užmiršimo vartams (angl. *Forget Gate*), $i_t$ - įvesties vartams (angl. *Input Gate*), $o_t$ - išvesties vartams (angl. *Output Gate*), taip pat $g_t$ - kandidatinei būsenai (angl. *Candidate State*) apskaičiuoti:

$$
\begin{aligned}
g_t = \tilde{C}_t &= \tanh(W_g x_t + U_g h_{t-1} + b_g) \\
f_t &= \sigma(W_f x_t + U_f h_{t-1} + b_f) \\
i_t &= \sigma(W_i x_t + U_i h_{t-1} + b_i) \\
o_t &= \sigma(W_o x_t + U_o h_{t-1} + b_o)
\end{aligned}
$$

Tuomet, naudojama praeito laiko momento atminties ląstelės būsena ($C_{t-1}$), kuri filtruojama užmiršimo vartų bei kandidatinė būsena, filtruojama įvesties vartų, apskaičiuojant einamuoju laiko momento atminties ląstelės būseną:

$$
C_t = g_t \circ i_t + f_t \circ C_{t-1}
$$

Galiausiai, normalizuojant būseną tanh funkcija bei filtruojant ją išvesties vartų pagalba, apskaičiuojama paslėpta būsena $h_t$, kuri iš principo yra išvestis, arba, esant sekančiam žingsniui, - jo įvestis:

$$
h_t = o_t \circ \tanh(C_t)
$$

## 1.3. LSTM panaudojimas nykstančio gradiento problemos sprendimui

Pagrindinis aspektas, padedantis LSTM modeliams spręsti nykstančio gradiento problemą yra minėtoji atminties ląstelių (kurių standartiniai RNN neturi) būsena. Dėl jos tiesiško veikimo principo gradientai nėra veikiami eksponentiškai juos keičiančių transformacijų, nuo kurių stipriai nukenčia RNN modelių mokymasis.

Kitas svarbus aspektas – vartų mechanizmas (kurio standartiniai RNN taip pat neturi), padedantys reguliuoti informacijos tėkmę į atminties ląstelės būseną bei iš jos:

1. Užmiršimo vartai – reguliuoja kiek informacijos iš praeitos ląstelės būsenos paliekama einamojoje
2. Įvesties vartai – reguliuoja, kiek naujos informacijos iš kandidatinės būsenos pridedama į atminties ląstelę
3. Išvesties vartai – reguliuoja kuri dalis atminties ląstelės įeis į išvestį.

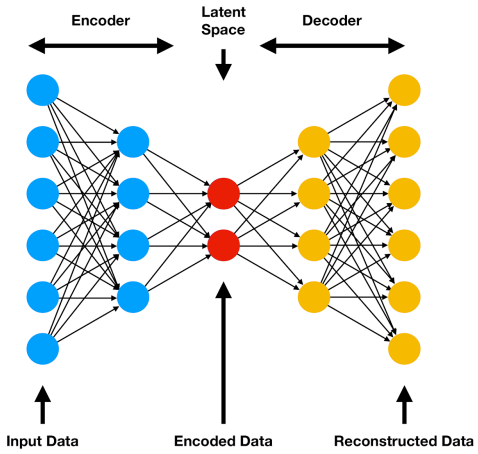
Taigi, apibendrinant, vartų mechanizmas, išmesdamas nereikalingą informaciją, padeda palaikyti adekvačią gradientų kaitą dirbant su ilgomis duomenų sekomis, kas pasekoje reiškia, kad yra palaikomas modelio mokymosi efektyvumas.

## 1.4. Kūrybiškumas

Kaip jau minėta, LSTM modelių stiprioji pusė – ta, jog jie geba atkurti ilgalaikius ryšius muzikoje – kas yra esminis dalykas norint palaikyti struktūrą. Tačiau išmokdami šiuos ryšius LSTM modeliai ne tik palaiko struktūrą, o ir geba sukurti kūrybinių variacijų [HS97].

# 2. VAE

VAE (angl. *Variational Autoencoders*) generavimui skirtų modelių tarpe išsiskiria unikaliu būdu modeliuoti duomenis. Skirtingai nuo kitų modelių, VAE modeliai užkoduoja ir išskirsto įvestis paslėptose erdvėse (angl. *latent spaces*) ir tada juos naudoja generuodami išvestis. Toks veikimo principas padeda šiems modeliams išmokti sudėtingas struktūras duomenyse, kas lemia VAE modelių naudojamumą muzikos generavime, turint galvoje sudėtingas muzikos struktūros ypatybes.




3 pav. VAE architektūra

Vienas iš didžiausių VAE modelių privalumų ir yra jų gebėjimas išmokti svarbias struktūras duomenyse. Muzikos generavimo kontekste tai reiškia gebėjimą užkoduoti esminius muzikos kompozicijos aspektus, tokius kaip ritmą, melodiją, harmoniją dinamiką ar toną. Todėl VAE modeliai gali išmokomi bendros, ilgalaikės muzikinių kūrinių struktūros [RER+18].

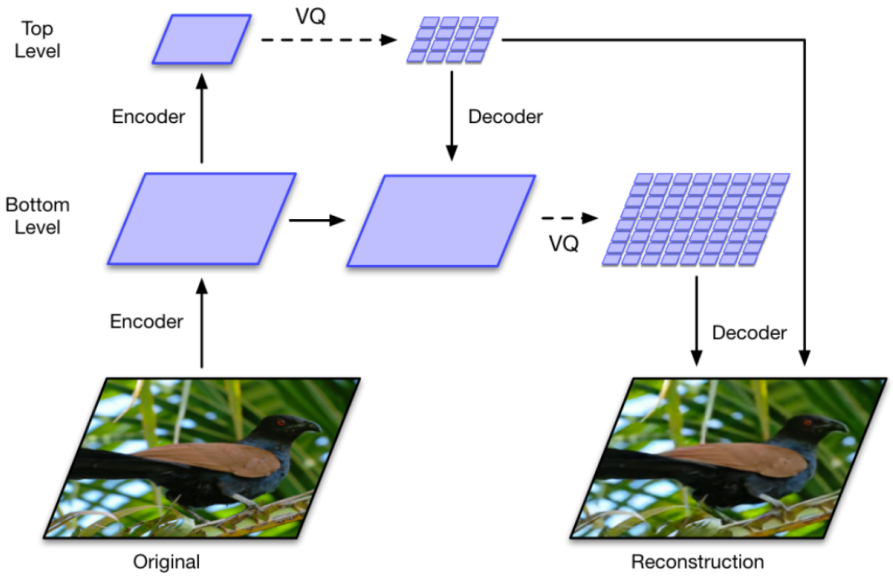
## 2.1. Kūrybiškumas

„Kūrybiškumą“ VAE modelių generuojamoje muzikoje lemia jų stochastinis veikimo principas. Jų duomenų modeliavimo principas leidžia generuoti muziką, kuri ne tik įvairi, tačiau ir pasižymi tam tikru nenuspėjamumu, kas yra vienas iš esminių meniško kūrybiškumo elementų. Šis darbas [BP20] iliustruoja, kaip VAE modeliai gali būti naudojami generuojant naujovišką muziką neprarandant mokymo duomenų esmės, tokiu būdu palaikant pusiausvyrą tarp novatoriškumo bei pažįstamų muzikos elementų.

## 2.2. VQ-VAE

VQ-VAE (angl. *Vector Quantized Variational Autoencoders*) - pažangesnė VAE modelių versija, kurioje minėtos paslėptos erdvės yra kvantuojamos pagal diskrečių reikšmių rinkinius. Jie pasižymi gebėjimu geriau koduoti struktūras - VQ-VAE modeliai koduoja, šiuo atveju, garso informaciją bei muzikines struktūras į paslėptus diskrečius vektorius, kuriais kompaktiškiau atspindima muzikos struktūra bei smulkesnės detalės.

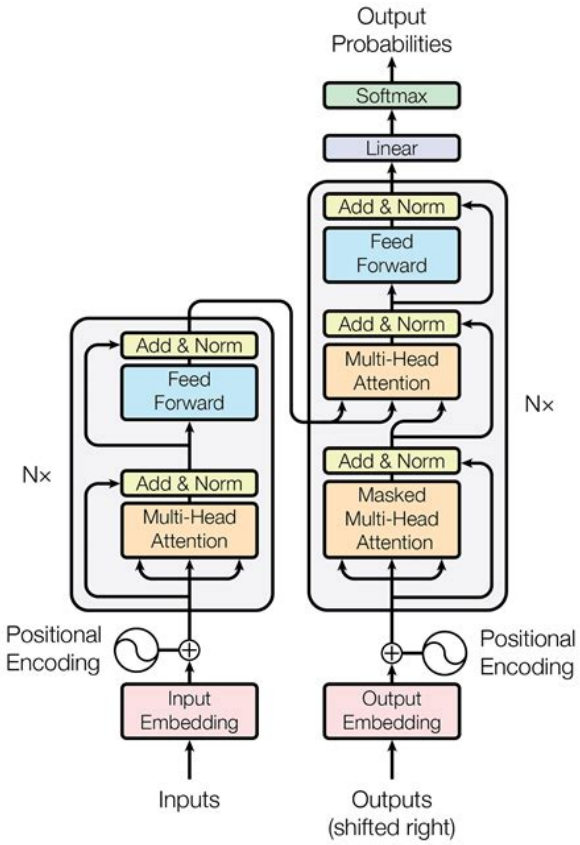
Esminis šių modelių bruožas jų hierarchiška struktūra - informacija koduojama dviem sluoksniais - už smulkesnes detales, kaip, pavyzdžiui, natos tembras ar garsumas, atsakingas žemesnis sluoksnis, o už stambesnes, kaip muzikinio kūrinio bendra struktūra - aukštesnis. Tai jiems leidžia gerai atkartoti bendrą muzikos struktūrą, neprarandant gebėjimo kokybiškai perteikti detales [PLX+21].




4 pav. VQ-VAE hierarchiška struktūra

# 3. Transformeriai

Transformeriai (angl. *Transformers*) sukurti su tikslu efektyviau apdoroti ilgalaikes priklausomybes duotos sekos duomenyse ir šis gebėjimas - vienas iš didžiausių jų privalumų [**vaswani2017attention**]. Muzikos generavimo kontekste tai reiškia gebėjimą palaikyti nuoseklią muzikos struktūrą.




5 pav. Transformerio architektūra

Šio transformerių gebėjimo priežastis - „dėmesio sau“ (angl. *self-attention*) mechanizmas, esantis šių modelių veikimo pagrindas. Jis leidžia sukurti ryšius tarp bet kurių elementų duomenų sekoje. Tai pasiekiama kiekvienam elementui priskiriant reikšmes, kurie nusako kitų sekos elementų svarbą to elemento atžvilgiu

Tai leidžia transformeriams vertinti skirtingų įvesties sekos dalių svarbą ir jie yra itin naudingi atliekant užduotis, reikalaujančias suprasti sudėtingus duomenų santykius. Tuo ir atsiskleidžia jų aktualumas muzikos generavime - šiuo atveju kalbant apie sudėtingų muzikinių struktūrų išmokimą ir atkartojimą.

## 3.1. Kūrybiškumas

Transformeriai architektūra iš prigimties palaiko kūrybiškumą muzikos generavime būtent dėl jų gebėjimo užfiksuoti ir atkartoti įvairius duomenų santykius. Minėtasis dėmesio sau mecha-

nizmas leidžia transformeriams generuoti kūrinius su sudėtingomis ir unikaliomis struktūromis bei ilgesnėmis melodijomis neprarandant nuoseklumo.

# 4. Kokybės vertinimas

## 4.1. Metodai

Dirbtinio intelekto generuojamos muzikos vertinimui pasirinktos kelios toliau įvardinamos metrikos, pagal kurias vertinami audio failai. Audio failų pavyzdžiai generuoti keliais žinomais modeliais:

1. „OpenAI“ Jukebox [DJP+20] - VQ-VAE tipo modelis.
2. „OpenAI“ Musenet [Pay19] - transformerio tipo modelis.
3. „Google“ projekto „Magenta“ [Eck16] modelis MusicRNN - LSTM tipo modelis.

Audio failų duomenys surinkti bei iš jų diagramos generuotos Python kalbos kodo pagalba.

## 4.2. Analizė

Toliau bus pateikiami duomenys diagramų formoje, kurių pagalba lyginami audio failai. Muzika buvo vertinama pagal tonus, ritmą, dinamiką bei tembrą. Atliekami ir lygiagrečiai aprašomi du tyrimai.

Pirmame modelių buvo generuojama klasikinė muzika. Todėl objektyvaus palyginimo tikslu taip pat vertinamas originalus, būtent klasikinės muzikos, kūrinio - Johano Sebastiano Bacho „Sinfonia, Oratorio BWV 249 Philippe Herreweghe” - ištrauka.

Diagramose:

1. 1-as failas („File 1“) : Iškarpa iš J. S. Bacho kūrinio „Sinfonia, Oratorio BWV 249 Philippe Herreweghe”.
2. 2-as failas („File 2“): „OpenAI“ Jukebox (VQ-VAE) modelio generuotos klasikinės muzikos iškarpa .
3. 3-ias failas („File 3“): „OpenAI“ MuseNet (transformerio) modelio generuotos klasikinės muzikos iškarpa.
4. 4-as failas („File 4“): „Google“ Magenta MusicRNN (LSTM) modelio generuotos klasikinės muzikos iškarpa.

Antrame eksperimente generuojama pop muzika. Palyginimui vertinama atlikėjos Lady Gaga kūrinio „Just Dance” instrumentinės versijos ištrauka.

Diagramose:

1. 1-as failas („File 1“) : Iškarpa iš atlikėjos Lagy Gaga kūrinio „Just Dance” instrumentinės versijos.
2. 2-as failas („File 2“): „OpenAI“ Jukebox (VQ-VAE) modelio generuotos pop muzikos iškarpa .
3. 3-ias failas („File 3“): „OpenAI“ MuseNet (transformerio) modelio generuotos pop muzikos iškarpa.
4. 4-as failas („File 4“): „Google“ „Magenta“ MusicRNN (LSTM) modelio generuotos pop muzikos iškarpa.

### 4.2.1. Entropija

Entropija - statistinio neapibrėžtumo (netikėtumo) matas, kuris yra naudojamas įvairiose srityse. Muzikos kontekste entropijos pagalba galima nusakyti kūrinio unikalumą, kuris atsispindi skirtinguose muzikos aspektuose, kaip tonas, ritmas, dinamika ir tembras.

Muzikos analizei dažnai naudojama Šenono entropija [Sha48].

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i)$$

kur:

- $X$ yra atsitiktinis kintamasis su $n$ galimų baigčių.
- $p(x_i)$ yra tikimybė, kad įvyks kiekviena unikali vertė $x_i$ duomenų rinkinyje.

### 4.2.2. Tonas

Tonų pasiskirstymo diagramos rodo, kaip dažnai skirtingi tonai pasitaiko analizuojamuose failuose. Vertikalioje (Y) ašyje pateikiama vertė rodanti, kiek kartų tam tikras tonas sutinkamas tam tikrame faile. Horizontalioji (X) ašis atvaizduoja tono vertę. Šiuo atveju naudojamas matavimo vienetas yra MIDI numeris (angl. *MIDI number*) - vienetas, kurio reikšmės atitinka tam tikrus natų tonus pagal MIDI standartą.

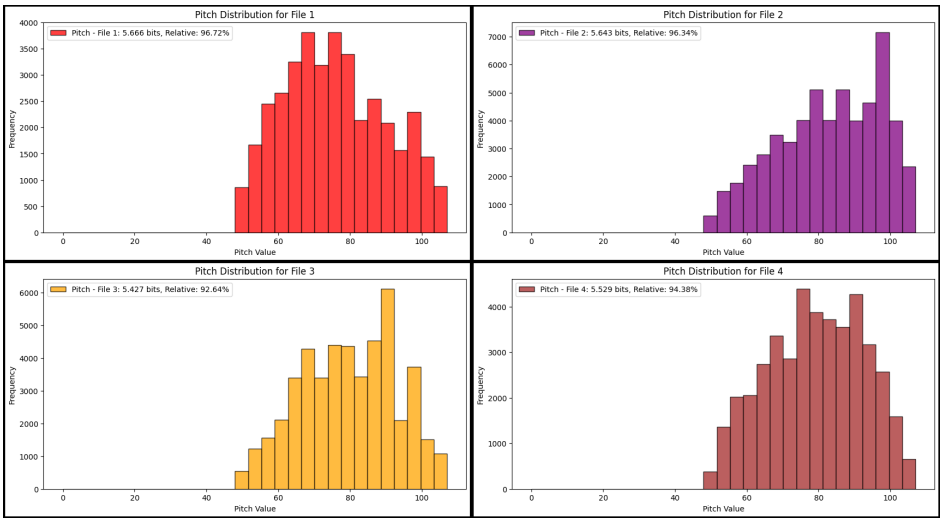
| Octave number | Note number | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | C | Db | D | Eb | E | F | Gb | G | Ab | A |
| -1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 0 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 |
| 1 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 |
| 2 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 |
| 3 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 |
| 4 | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 |
| 5 | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 | 80 | 81 |
| 6 | 84 | 85 | 86 | 87 | 88 | 89 | 90 | 91 | 92 | 93 |
| 7 | 96 | 97 | 98 | 99 | 100 | 101 | 102 | 103 | 104 | 105 |
| 8 | 108 | 109 | 110 | 111 | 112 | 113 | 114 | 115 | 116 | 117 |
| 9 | 120 | 121 | 122 | 123 | 124 | 125 | 126 | 127 | | |

6 pav. MIDI numerių reikšmės

Svarbu paminėti, jog:

- žemais tonais laikomi tie, kurių MIDI numerio reikšmės yra 0-40,
- vidutiniais - 41-80,
- aukštais - 81-127.

#### 4.2.2.1. Pirmasis eksperimentas

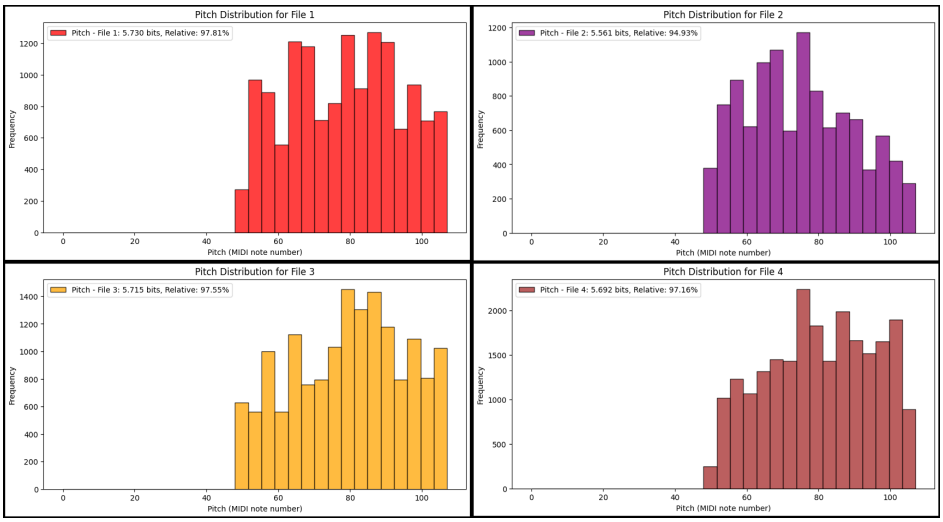



7 pav. Tonų diagrama pirmajame eksperimente

Diagramų rezultatai:

1. File 1: Bacho kūrinyje tonų pasiskirstymas yra ganėtinai subalansuotas. Entropijos vertė 5.666 bitų (96.72%). Tai reiškia, jog kūrinys turi daug įvairių dažnai pasikartojančių tonų, kurie sudaro sudėtingą muzikinę struktūrą, kas rodo tikrai aukštą toninės struktūros unikalumą.
2. File 2: VQ-VAE modelio generuota muzika taip pat rodo aukštą entropiją (5.643 bitų, 96.34%), kas taip pat reiškia gerą balansą bei aukštą unikalumo lygį, tačiau matomas didesnis akcentas aukštesniuose tonuose.
3. File 3: Transformerio išvestis rodo šiek tiek mažesnę entropiją (5.427 bitų, 92.64%). Vadinasi, tonai šiame kūrinyje yra kiek labiau nuspėjami. Taip pat matoma tendencija į kiek aukštesnius tonus.
4. File 4: LSTM modelio generuota muzika rodo, palyginus, vidutinę, tačiau taip pat aukštą entropiją (5.529 bitų, 94.38%). Pats tonų pasiskirstymas - panašiausias į J. S. Bacho kūrinį ir geriausiai subalansuotas.

#### 4.2.2.2. Antrasis eksperimentas




8 pav. Tonų diagrama antrajame eksperimente

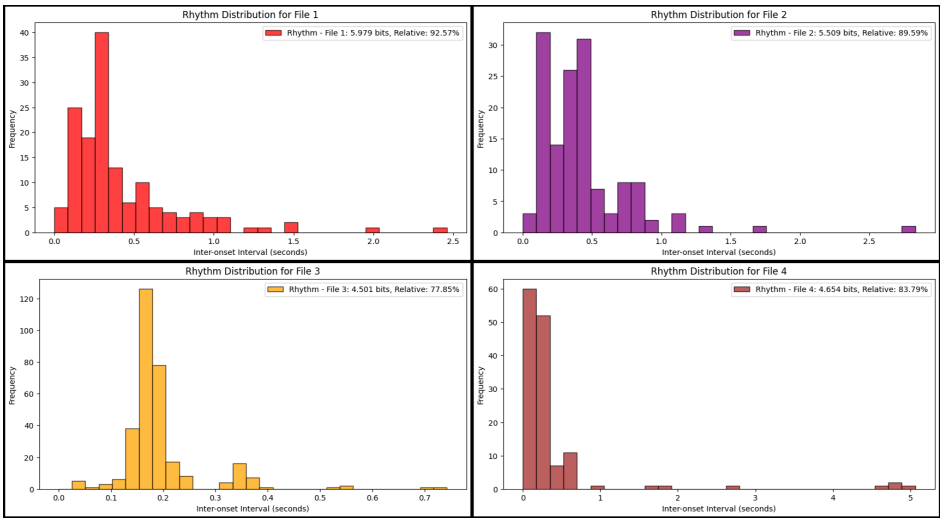
Diagramų rezultatai:

1. File 1: Lady Gaga kūrinyje tonų pasiskirstymas yra turi gerą balansą. Entropijos vertė 5.730 bitų (97.81%). Tai reiškia, jog šis kūrinys, kaip ir J. S. Bacho, turi plačią dažnai pasikartojančių tonų įvairovę. Čia taip pat matomas aukštas toninės struktūros unikalumas.
2. File 2: VQ-VAE modelio generuota muzika rodo mažesnę, nors vis tiek pakankamai aukštą entropiją (5.561 bitai, 96.34%), irgi reiškiančią neblogą tonų balansą bei aukštą, nors ir kiek žemesnį unikalumo lygį. Bendras tonų pasiskirstymas taip pat šiek tiek kitoks - matomas polinkis į kiek žemesnius tonus.
3. File 3: Transformerio išvestis rodo aukštą entropiją, beveik siekiančią Lady Gaga kūrinio entropiją. (5.715 bitai, 97.55%). Vadinasi, tonai šiame kūrinyje turi didelį nenuspėjamumą. Bendras tonų pasiskirstymas taip pat labai panašus į Lady Gaga kūrinį.
4. File 4: LSTM modelio generuota muzika taip pat rodo aukštą entropiją, vos atsiliekančią nuo Lady Gaga kūrinio bei transformerio išvesties (5.692 bitai, 97.16%). Bendras tonų pasiskirstymas taip pat ganėtinai panašus į Lady Gaga kūrinį.

### 4.2.3. Ritmas

Ritmo diagramos rodo skirtingų intervalų tarp natų (angl. *interonset interval*) pasiskirstymą. Vertikalioje (Y) ašyje pateikiama vertė, rodanti, kiek kartų tam tikras intervalas pasitaiko audio faile. Horizontali (X) ašis rodo intervalo trukmę sekundėmis.

#### 4.2.3.1. Pirmasis eksperimentas

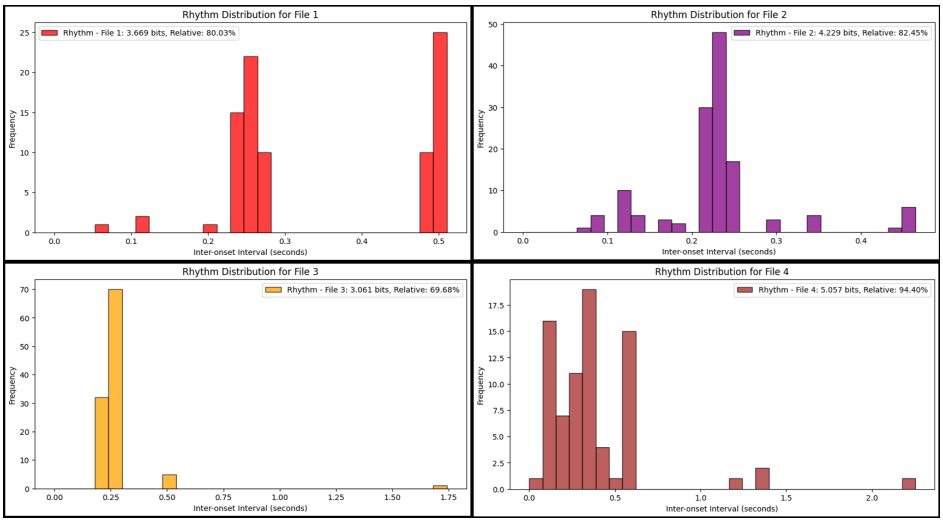



9 pav. Ritmo diagrama pirmajame eksperimente

Diagramų rezultatai:

1. File 1: Bacho kūrinyje šie intervalai yra ganėtinai įvairūs. Taip pat matoma aukšta entropija (5.979 bitai, 92.57%) reiškianti unikalią ritminę struktūrą.
2. File 2: VQ-VAE modelis rodo panašų intervalų pasiskirstymą, tačiau kiek mažesnę entropiją (5.509 bitai, 89.95%). Tai reiškia, kad modelis geba atkartoti muzikos ritminę struktūrą bei turi pakankamai sudėtingą ritminę struktūrą.
3. File 3: Transformerio išvesties ritminė struktūra yra labiau koncentruota ties trumpesniais intervalais. Taip pat matoma mažiausia entropija (4.501 bitai, 77.85%) Tai reiškia, jog šis modelis generuoja ritmą su trumpesniais, labiau nuspėjamais ir pasikartojančiais intervalais tarp natų.
4. File 4: LSTM modelis taip pat įvertintas kiek mažesne entropija (4.654 bitai, 83.79%), taip pat tačiau turi plačiausią intervalų tarp natų diapazoną. Tai, autoriaus nuomone, reiškia, jog šis modelis prasčiausiai atkartoja muzikos ritminę struktūrą.

#### 4.2.3.2. Antrasis eksperimentas




10 pav. Ritmo diagrama antrajame eksperimente

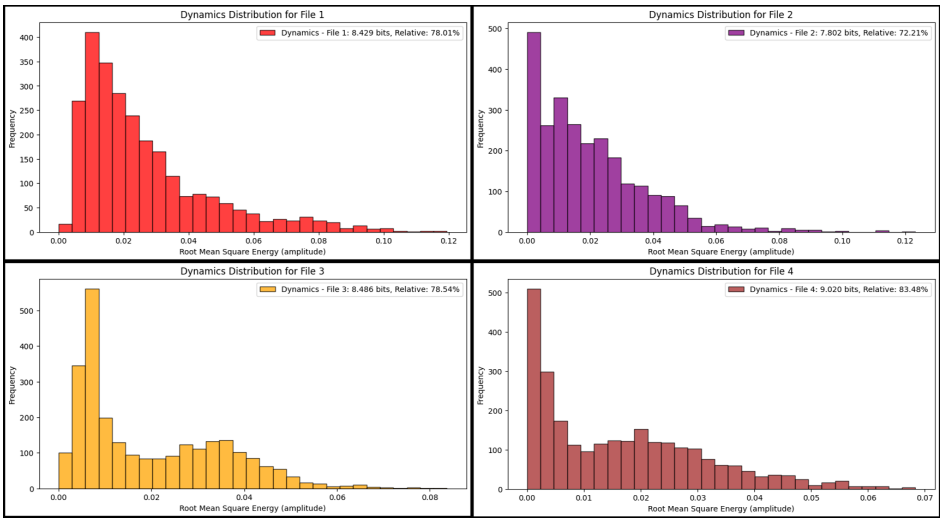
1. File 1: Lady Gaga kūrinyje šie intervalai nėra įvairūs. Taip pat matoma visai žema entropija (3.669 bito, 80.03%). Tai reiškia paprastą, nuspėjamą ritminę struktūrą.
2. File 2: VQ-VAE modelis rodo panašų intervalų diapazoną, tačiau su daugiau įvairovės. Taip pat matoma kiek aukštesnė entropiją (4.229 bito, 82.45%). Tai reiškia, kad modelio išvestis turi sudėtingesnę muzikos ritminę struktūrą su daugiau unikalumo. Taip pat matomas didesnis polinkis į trumpesnius intervalus.
3. File 3: Transformerio išvestis yra itin koncentruota ties trumpesniais intervalais. Visiškai nėra įvairovės ritminėje struktūroje. Taip pat matoma mažiausia entropija (3.061 bito, 69.68%). Tai reiškia, jog šis modelis generuoja ritmą su trumpesniais, labiau nuspėjamais ir pasikartojančiais intervalais tarp natų.
4. File 4: LSTM modelio išvestis šiuo atveju rodo aukščiausią entropiją (5.057 bitai, 83.79%). Tai reiškia didžiausią unikalumą. Taip pat matomas stiprus polinkis į gerokai trumpesnius intervalus.

### 4.2.4. Dinamika

Dinamikos diagramos rodo skirtingų garso stiprumo lygių pasiskirstymą. Vertikalioje ašyje (Y) pateikiama vertė, rodanti, kiek kartų tam tikras garso stiprumo lygis pasitaiko audio faile. Horizontali ašis (X) rodo stiprumo lygį, matuojamą RMS (angl. *Root Mean Square*).

RMS - vienetas, skirtas matuoti signalo stiprumą per tam tikrą laiko tarpą. Muzikos kontekste RMS rodo paprasčiausiai garsumą.

#### 4.2.4.1. Pirmasis eksperimentas

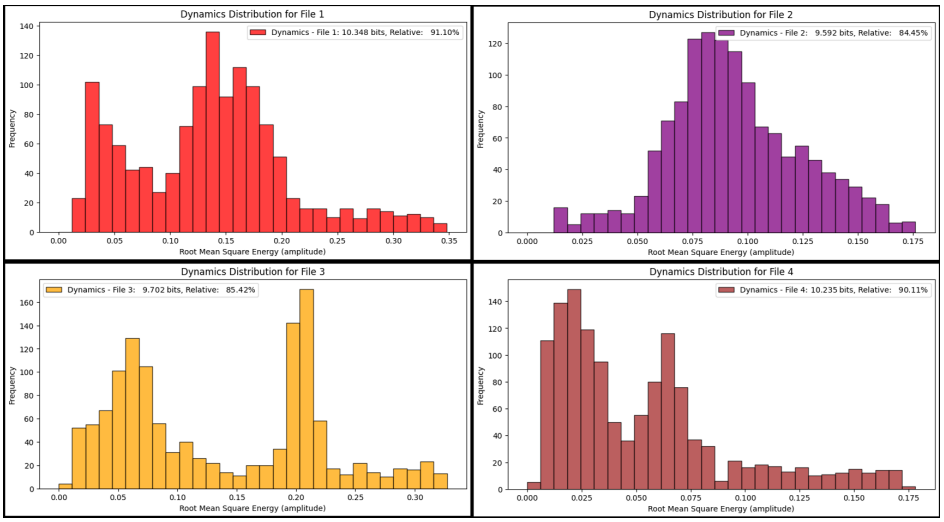



11 pav. Dinamikos diagrama pirmajame eksperimente

Diagramų rezultatai:

1. File 1: J. S. Bacho kūrinyje matomas ganėtinai netolygus garsumo pasiskirstymas su dideliu polinkiu į tylesnius garsus. Taip pat matoma žema entropija (8.429 bito, 78.01%). Tai mažą garsumo varijavimą.
2. File 2: VQ-VAE modelio išvestyje matomas panašus garsumo diapazonas, su polinkiu į dar tylesnius garsus. Taip pat matoma dar žemesnė entropija (7.802 bito, 72.21%). Tai reiškia, kad modelio išvestis turi dar labiau nuspėjamą garsumo visumą.
3. File 3: Transformerio išvesties garsumo diapazonas siauresnis, nei J. S. Bacho kūrinio, tačiau turi stipriausią polinkį panašioje vietoje. Taip pat matoma ir panaši entropija (8.486 bito, 78.54%). Vadinasi, šio modelio išvesties garsumo visuma turi panašaus, žemo lygio nuspėjamumą, kaip ir J. S. Bacho kūrinio.
4. File 4: LSTM modelio išvestis rodo šiuo atveju aukščiausią entropiją (9.020 bitų, 83.48%). Tai reiškia didžiausias variacijas garsume. Tačiau matomas polinkis į tylesnius garsus, nei J. S. Bacho kūrinyje, panašus į VQ-VAE modelio.

#### 4.2.4.2. Antrasis eksperimentas




12 pav. Dinamikos diagrama antrajame eksperimente

Diagramų rezultatai:

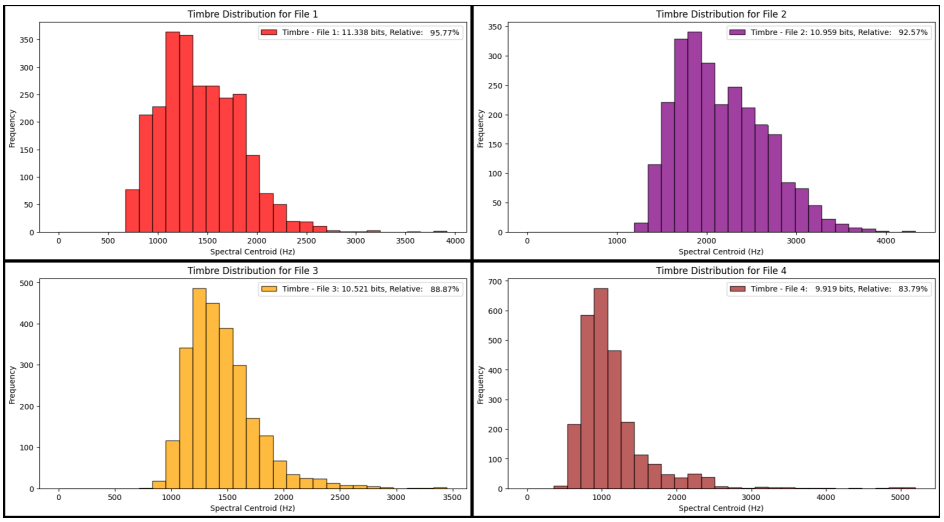
1. File 1: Lady Gaga kūrinyje matomas tolygiausias garsumo pasiskirstymas. Taip pat matoma aukščiausia entropija (10.348 bito, 91.10%). Tai reiškia didžiausią originalumą garsumo variacijų naudojime.
2. File 2: VQ-VAE modelis rodo mažesnį garsumo diapazoną, su didesniu polinkiu į tylesnius garsus. Taip pat matoma žemesnė entropija (9.592 bito, 84.45%). Tai reiškia, kad modelio išvestis turi labiau nuspėjamą garsumo visumą.
3. File 3: Transformerio išvesties garsumo diapazonas labai panašus į Lady Gaga kūrinio. Tačiau, matoma mažesnė entropija, panaši į VQ-VAE modelio (9.702 bito, 85.42%). Vadinasi, šio modelio garsumo visuma taip pat labiau nuspėjama.
4. File 4: LSTM modelio išvestis rodo aukštą entropiją, beveik tokią pat, kaip Lady Gaga kūrinio (10.057 bitų, 90.11%). Tai reiškia didelį unikalumą garsumo variacijose. Tačiau matomas polinkis į tylesnius garsus, panašus į VQ-VAE modelio.

### 4.2.5. Tembras

Tembro diagramos rodo skirtingų tembro charakteristikų pasiskirstymą. Vertikalioje (Y) ašyje pateikiama dažnio vertė, rodanti, kiek kartų tam tikras tembro spektrinis centroidas (angl. *spectral centroid*) pasitaiko audio faile. Horizontali (X) ašis rodo spektrinio centroido reikšmę.

Spektrinis centroidas yra dažnis, kuris atspindi energijos pasiskirstymo vidurkį. Iš esmės, juo pateikiama viena vertė, nusakanti, kur (šiuo atveju, kokiame garso dažnyje), koncentruota daugiausia energijos.

#### 4.2.5.1. Pirmasis eksperimentas




13 pav. Tembro diagrama pirmajame eksperimente

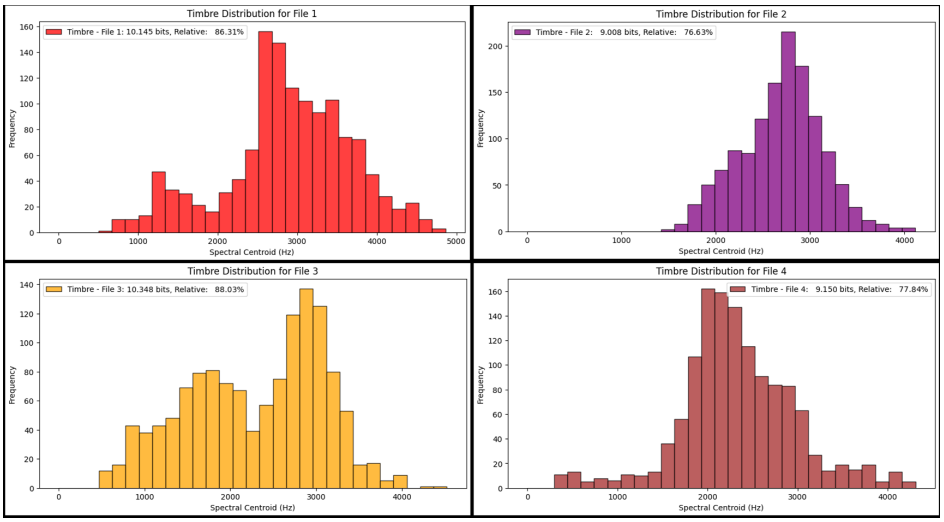
Diagramų rezultatai:

Visuose failuose matomas ganėtinai panašus tembro diapazonas, ganėtinai tiksliai atitinkantis J. S. Bacho kūrinio, išskyrus VQ-VAE, kurios tembro vidurkis akivaizdžiai aukštesnis.

Entropijos:

1. File 1: Bacho kūrinyje aukščiausia (11.338 bitai, 95.77%). Tai rodo didžiausias variacijas tembre.
2. File 2: VQ-VAE modelio generuoto kūrinyje matoma kiek mažesnė entropija (10.959 bitai, 92.57%). Tai rodo panašias, taip pat ganėtinai aukšto lygio tembro variacijas.
3. File 3: Transformerio išvesties entropija - dar mažesnė (10.521 bitai, 88.87%). Matomos jau stipresnės tendencijos ties tam tikrais dažniais, reiškiančios mažesnio variacijas tembre.
4. File 4: LSTM modelio išvesties entropija - pati mažiausia (9.919 bitai, 83.79%). Tai reiškia ganėtinai pastovų, mažai varijuojantį tembrą. Taip pat matoma tendencija į šiek tiek mažesnius dažnius, nei J. S. Bacho kūrinyje.

#### 4.2.5.2. Antrasis eksperimentas




14 pav. Tembro diagrama antrajame eksperimente

Diagramų rezultatai:

1. File 1: Lady Gaga kūrinyje matoma viena iš aukščiausių, nors iš esmės ne itin aukšta entropija (10.145 bitų, 86.31%) Vadinasi, šiame kūrinyje nėra ypatingai varijuojantis tembras.
2. File 2: VQ-VAE modelis rodo panašų tembro diapazoną, su polinkiu į kiek aukštesnį. Taip pat matoma žemesnė entropija (9.008 bito, 76.63%). Tai reiškia, kad modelio išvestis turi dar labiau nuspėjamą, mažiau varijuojantį tembrą.
3. File 3: Transformerio išvesties tembro diapazonas taip pat panašus į Lady Gaga kūrinio, tačiau, kaip ir VQ-VAE modelio, rodo tendenciją į kiek aukštesnius dažnius. Šiuo atveju šio modelio išvesties entropija didžiausia, kiek didesnė ir už Lady gaga kūrinio (10.348 bito, 88.03%). Vadinasi, šio tembras labiausiai varijuoja.
4. File 4: LSTM modelio išvestis rodo žemą, panašią į VQ-VAE modelio išvesties entropiją (9.088 bitų, 76.63%). Tai reiškia žemą tembro variaciją. Tačiau matomas polinkis į žemesnius dažnius, panašius kaip Lady Gaga kūrinio.

## 4.3. Modelių vertinimai

### 4.3.1. VQ-VAE

- Aukšta tonų entropija, bei panašus į žinomus kūrinius tonų diapazonas abiejuose eksperimentuose.
- Patenkinama ritmo entropija antrame eksperimentuose, aukšta pirmame. Pirmame eksperimente bendros ritmo ypatybės yra taikliai atitinkančios J. S Bacho pavyzdį, tačiau antrajame eksperimente matomas nukrypimas.
- Dinamikos entropija, palyginus, žema abiejuose eksperimentuose. Diapazonas pirmame eksperimente panašus į pavyzdį, tačiau antrame - kur kas mažesnis.

- Tembro entropija pirmame eksperimente gana aukšta, antrame – pastebimai mažesnė. Diapazono atžvilgiu, matomi nukrypimai abiejuose eksperimentuose

Nenuspėjamumo, arba, kūrybiškumo, kontekste, modeliui puikiai sekėsi tonų bei ritmo atžvilgiu. Matomi sunkumai tembre ir ypač dinamikoje.

Kalbant apie panašumus į žmogaus kurtą muziką tiek tonai atitinka pavyzdį, tačiau kitais aspektais matomi nukrypimai, ypač tembre.

### 4.3.2. Transformeris

- Aukšta tonų entropija, bei panašus į žinomus kūrinius tonų diapazonas abiejuose eksperimentuose.
- Ritmo entropija žema abiejuose eksperimentuose. Bendrose ritmo ypatybėse matomi nukrypimai abiejuose eksperimentuose.
- Dinamikos entropija palyginus žema abiejuose eksperimentuose. Bendras diapazonas antrame eksperimente panašus į pavyzdį, tačiau pirmame matomas nukrypimas
- Tembro entropija pirmame eksperimente – ganėtinai žema, tačiau antrame – aukštesnė. Diapazonas panašus į pavyzdžius abiejuose eksperimentuose.

Kūrybiškumo kontekste, šiam modeliui geriausiai sekėsi su tonais. Tembro atžvilgiu pop kūrinio rezultatas – geras, klasikinės muzikos – prastas. Kitais aspektais matomas palyginus didelis nuspėjamumas.

Kalbant apie taiklumą žmogaus kurtos muzikos atžvilgiu, modelio išvestis yra panaši į pavyzdžius tonais bei tembru, tačiau matomi nukrypimai dinamikoje bei ypač ritme.

### 4.3.3. LSTM

- Aukšta tonų entropija, bei panašus į žinomus kūrinius tonų diapazonas abiejuose eksperimentuose.
- Ritmo entropija palyginus žema pirmame eksperimente, tačiau aukščiausia antrame. Bendros ritmo ypatybės nukrypusios nuo pavyzdžių abiejuose eksperimentuose
- Dinamikos entropija aukšta abiejuose eksperimentuose. Diapazonas siauresnis nei pavyzdžiai abiejuose eksperimentuose.
- Tembro entropija žema abiejuose eksperimentuose. Diapazonas abiejuose eksperimentuose panašus į pavyzdžius.

Kūrybiškumo kontekste, modelis demonstravo gerus rezultatus tonų bei dinamikos atžvilgiu. Ritmas buvo labiau nuspėjamas, tembras – labiausiai.

Kalbant apie panašumą į žmonių kurtus pavyzdžius, tonai bei tembras yra taiklūs. Nukrypimai matomi ritme bei dinamikoje.

## 4.4. Bendras Vertinimas

### 4.4.1. Tonas

- Visi modeliai rodo aukštą entropiją, kuri reiškia mažą nuspėjamumo, arba, aukštą kūrybiškumo laipsnį.
- Visų modelių generuotų kūrinių tonų diapazonas panašus į žmonių kurtus pavyzdžius - tai reiškia, jog visi modeliai taikliai atkartojo abiems žanrams būdingus tonus.

### 4.4.2. Ritmas

- Bendrai, visų modelių ritmo entropija yra žema abiejuose eksperimentuose. Išimtys - VQ-VAE aukšta entropija pirmame eksperimente, bei LSTM antrame.
- Ritmo ypatybės taip pat prastai atitinka žmonių kurtą muziką, išskyrus VQ-VAE modelio išvestį pirmame eksperimente.

### 4.4.3. Dinamika

- Visų modelių dinamikos entropija ganėtinai žema, išskyrus LSTM.
- Taiklumas lyginant su žmonių kurta muzika taip pat prastas, išimtys - VQ-VAE pirmame eksperimente bei transformeris antrame.

### 4.4.4. Tembras

- Bendrai, tembro entropijos pirmame eksperimente gerokai aukštesnės, nei antrame - įskaitant ir žmonių kurtą muzika
- Abiejuose eksperimentuose diapazonai panašūs į pavyzdžius, išskyrus VQ-VAE išvestis.

# 5. Integracija į Blockly

## 5.1. Modelio pasirinkimas

Autoriaus nuomone, integracijai į Blockly iš aptartų modelių labiausiai tiktų Jukebox. Taip yra todėl, jog jis leidžia tiek generuoti naujas, tiek tęsti esamas melodijas. Taip pat jis turi paprastą pasirinkimą tarp įvairių skirtingų muzikos žanrų.

Praeitame pusmetyje autoriaus buvo pradėtas projektas, kuriame vartotojui leidžiama kurti muzikines kompozicijas naudojant Blockly sąsają bei Tone.js JavaScript kalbos biblioteką.

Modelis turėtų būti užkraunamas tuo metu, kai paleidžiama programa, ant atskiros gijos (angl. *thread*) – tai leis naudotojui atlikti kitus veiksmus, kol modelis generuos išvestis.


15 pav. Pradėtas projektas, leidžiantis kurti muzikines kompozicijas Blockly vizualinės sąsajos pagalba.

## 5.2. Integracija

### 5.2.1. Vartotojo sąsaja

Turėtų būti aprašomi naujo blokai:

- Esamos melodijos tęsimui.
- Naujos melodijos kūrimui.

Šie blokai turėtų turėti žanro pasirinkimą, būdingą Jukebox sąsajai.

### 5.2.2. Veikimo procesas

Blockly pagrindu veikiančios programos bei Jukebox modelio sąveika galėtų atrodyti taip:

1. Jei tęsiama melodija, ji konvertuojama į modeliui tinkamą garso formatą.
2. Modelis sugeneruoja išvestį.

3. Išvestis konvertuojama į MIDI formatą, iš kurio tada yra perteikiama natų bei pauzių blokais Blockly pagrindu veikiančios programos vartotojo sąsajoje.
4. Melodija pateikiama vartotojui pažįstamu formatu ir gali būti išklausoma naudojant vartotojo sąsają.

## 5.3. Aktualumas

Toks įrankis iš esmės apjungtų autoriaus pradėtą projektą leidžiantį kurti kompozicijas naudojant Blockly bei Jukebox galimybes generuoti melodijas, bei tęsti jau esamas. Tai galėtų supaprastinti procesą, pavyzdžiui, sukūrus melodiją Blockly pagrindo programoje ir norint ją pratęsti naudojant dirbtinį intelektą. Tačiau, autoriaus nuomone, paklausa bei reikmė tokios situacijos paprastinimui nėra reikšminga, todėl projekto nuspręsta neįgyvendinti.

# Rezultatai

1. Išanalizuotos trijų tipų modelių (LSTM, VQ-VAE, transformerių) galimybės generuojant muziką.
2. Pristatytas būdas vertinti generuojamos muzikos kokybę.
3. Atlikti du eksperimentai, kurių metu kiekvienu iš minėtų modelių buvo generuoti klasikinės bei pop muzikos kūriniai, bei įvertinta jų kokybė kūrybiškumo atžvilgiu, taip pat kaip jie atitinka žmonių kurtos muzikos pavyzdžius.
4. Pasiūlytas dirbtinio intelekto generuojamos muzikos integracijos į Blockly pagrindu sukurtą programą konceptas.

# Išvados

1. Dirbtinio intelekto generuojamą muziką kokybiškai galima vertinti pasitelkiant muzikines metrikas, tokias kaip tonai, ritmas, tembras, ar dinamika ir jas lyginant tarpusavyje bei su žmogaus kurtos muzikos pavyzdžiu.
2. Galima išskirti tam tikras aiškias modelių stiprybes:
   (a) VQ-VAE:
      i. Aukštas kūrybiškumo lygis tonų bei ritmo atžvilgiu.
      ii. Taikliai žmonių kuriamą muziką atitinkantys tonai.
   (b) Transformeriai:
      i. Aukštas kūrybiškumo lygis tonų atžvilgiu.
      ii. Taikliai žmonių kuriamą muziką atitinkantys tonai bei tembras.
   (c) LSTM:
      i. Aukštas kūrybiškumo lygis tonų bei dinamikos atžvilgiu.
      ii. Taikliai žmonių kuriamą muziką atitinkantys tonai bei tembras.
3. Remiantis vertinimo duomenimis, VQ-VAE modelis demonstravo didžiausią bendrą unikalumą bei gebėjimą teisingai atkartoti generuojamos muzikos ritmą.
4. Dirbtinio intelekto modelio generuojamos muzikos unikalumas, arba, modelio „kūrybiškumas“ kai kuriais atvejais gali reikšti nukrypimą nuo teisingų bandomos atkartoti muzikos savybių.
5. Dirbtinio intelekto generuojamą muziką galima integruoti į Blockly pagrindu veikiančią programą, tačiau tam nėra reikmės.

# Šaltiniai


[BP20] J.-P. Briot, F. Pachet. Deep learning for music generation: challenges and directions. *Neural Computing and Applications*. 2020, tomas 32, numeris 4, p. 981–993.

[BSF94] Y. Bengio, P. Simard, P. Frasconi. Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*. 1994, tomas 5, numeris 2, p. 157–166.

[CCC+22] M. Civit, J. Civit-Masot, F. Cuadrado, M. J. Escalona. A systematic review of artificial intelligence-based music generation: Scope, applications, and future trends. *Expert Systems with Applications*. 2022, p. 118190.

[DJP+20] P. Dhariwal, H. Jun, C. Payne, J. W. Kim, A. Radford, I. Sutskever. Jukebox: A generative model for music. *arXiv preprint arXiv:2005.00341*. 2020.

[Eck16] D. Eck. *Magenta*. 2016. Prieiga per internetą: `magenta.tensorflow.org/blog/2016/06/01/welcome-to-magenta/`.

[HB22] C. Hernandez-Olivan, J. R. Beltran. Music composition with deep learning: A review. *Advances in speech and music technology: computational aspects and applications*. 2022, p. 25–50.

[HS97] S. Hochreiter, J. Schmidhuber. Long short-term memory. *Neural computation*. 1997, tomas 9, numeris 8, p. 1735–1780.

[KFV20] M. Kaliakatsos-Papakostas, A. Floros, M. N. Vrahatis. Artificial intelligence methods for music generation: a review and future perspectives. *Nature-Inspired Computation and Swarm Intelligence*. 2020, p. 217–245.

[LGL+21] X. Liu, K. Gao, B. Liu, C. Pan ir kiti. Advances in deep learning-based medical image analysis. *Health Data Sci*. 2021, tomas 2021, p. 8786793.

[Pay19] C. Payne. *MuseNet*. 2019. Prieiga per internetą: `openai.com/blog/musenet`.

[PLX+21] J. Peng, D. Liu, S. Xu, H. Li. Generating diverse structure for image inpainting with hierarchical VQ-VAE. Iš: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, p. 10775–10784.

[RER+18] A. Roberts, J. Engel, C. Raffel, C. Hawthorne, D. Eck. A hierarchical latent vector model for learning long-term structure in music. Iš: *International conference on machine learning*. PMLR, 2018, p. 4364–4373.

[Sha48] C. E. Shannon. A mathematical theory of communication. *The Bell system technical journal*. 1948, tomas 27, numeris 3, p. 379–423.

[SJ19] K. Smagulova, A. P. James. A survey on LSTM memristive neural network architectures and applications. *The European Physical Journal Special Topics*. 2019, tomas 228, numeris 10, p. 2313–2324.

[XWY+23] Z. Xiong, W. Wang, J. Yu, Y. Lin, Z. Wang. A Comprehensive Survey for Evaluation Methodologies of AI-Generated Music. *arXiv preprint arXiv:2308.13736*. 2023.

[ZA21] K. Zhang, A. B. Aslan. AI technologies for education: Recent research & future directions. *Computers and Education: Artificial Intelligence*. 2021, tomas 2, p. 100025.