

TARTU ÜLIKOOL  
Arvutiteaduse instituut  
Infotehnoloogia mitteinformaatikutele õppekava

**Raivo Kasepuu**  
**Muusika toonimise kasutamine muusika žan-**  
**rite klassifitseerimise mudelites**  
**Magistritöö (15 EAP)**

Juhendaja: Anna Aljanaki, PhD

Tartu 2023

## **Muusika toonimise kasutamine muusika žanrite klassifitseerimise mudelites**

### **Lühikokkuvõte:**

Käesoleva magistritöö raames teostatakse uurimus leidmaks, kuidas muusikafailide toonimised mõjuvad muusika žanrite klassifitseerimiste mudelite täpsusele. Püstitatud ülesande lahendamiseks võrreldakse toonimata andmestikuga etalon mudeli täpsust erinevate toonitud andmestike abil loodud mudelite täpsustega. Töö sisendiks on GTZAN muusika andmestik ja muusika toonimisi uuritakse MFCC koefitsientidel põhinevatel muusika žanrite klassifitseerimise mudelitel. Töö tulemusel selgus, et toonitud muusikaga rikastatud andmestikel treenitud muusika žanri klassifitseerimise mudelid on keskmiselt täpsemad kui ainult toonimata muusika muusikal treenitud mudelid.

### **Võtmesõnad:**

Andmeanalüüs, andmestike rikastamine, muusika toonimine, GTZAN, MFCC

### **CERCS:**

P176 Tehisintellekt

## **Music augmentation usage in music genre classification models**

### **Abstract:**

In this master's thesis, research is carried out to find out how the music augmentation affects the accuracy of music genre classification models. To solve the posed problem, the accuracy of the standard model with the original non-augmented dataset is compared with the accuracy of the models created using different augmented datasets. The master thesis uses GTZAN music dataset and music augmentations are observed on MFCC coefficients-based music genre classification models. As a result of the research, it was found that music genre classification models trained on augmented music datasets are more accurate than models trained on non-augmented music datasets.

### **Keywords:**

Data analysis, data augmentation, music augmentation, GTZAN, MFCC

### **CERCS:**

P176 Artificial intelligence

## Sisukord

1.	Sissejuhatus .....	4
2.	Mõisted ja terminid.....	5
3.	Taustainfo .....	6
3.1	MIR - muusika andmeanalüütika .....	6
3.2	Muusika žanrite automaatne klassifitseerimine.....	6
3.3	MFCC kasutamine.....	7
3.4	Muusika andmestikud. <i>GTZAN</i> andmestik.....	11
3.5	Muusika andmestike rikastamine .....	11
3.6	Muusikafailide toonimine.....	11
3.7	Varasemad uuringud.....	18
4.	Metoodika.....	20
4.1	Töö andmestikuga. ....	20
4.2	Mudeli optimeerimine ja treenimine <i>GTZAN</i> originaalandmestikul .....	21
4.3	Erinevate toonimiste mõjude võrdlemine.....	21
5.	Tulemused .....	24
5.1	Erinevate toonimiste meetoditega saavutatud mudelite täpsused.....	24
5.2	Tulemuste analüüs. ....	25
6.	Kokkuvõte .....	26
7.	Viidatud kirjandus .....	27
Lisad	.....	29
I.	Tulemuste tabelid, graafikud ja segadusmaatriksid.....	29
II.	Kasutatud tarkvara lähtekoodid .....	35
III.	Litsents .....	36

## 1. Sissejuhatus

Muusika žanrite automaatne klassifitseerimine on muusika informaatika valdkonnas oluline uurimisteema, mis aitab kaasa žanripõhiste soovitusüsteemide, muusikaotsingute ja digitaalse arhiveerimise arendamisele.

Viimastel aastatel on sügavõppe meetodite kasutamine žanri klassifitseerimise täpsuse parandamisel näidanud märkimisväärset edu. Üks võimalus täpsuse suurendamiseks on kasutada muusika andmete rikastamist (ingl *music data augmentation*) mis hõlmab olemasolevate andmekogumite laiendamist kasutades muusika toonimist (ingl *music augmentation*), lisades neile erinevate toonimismeetoditega modifitseeritud versioone algsest sisust [10].

Andmete rikastamise peamine eesmärk on luua mitmekesisem andmestik, mis võimaldab masinõppemudelitel saada paremaid ja üldistatavamaid tunnuseid. Muusika žanrite klassifitseerimise kontekstis võib muusika toonimine hõlmata erinevaid heli modifitseerimise tehnikaid - nagu heli pikkuse muutmine, helitugevuse muutmine, müra lisamine või eemaldamine ja heli segamine teiste helidega.

Erinevate toonimismeetodite rakendamine muusikažanrite klassifitseerimise mudelitele võivad suurendada nende täpsust. Varasemad uuringud on näidanud, et mõned toonimise meetodid, nagu helistiku muutmine, helitugevuse muutmine heli spektri erinevates lõikudes annavad parema täpsuse kui teised [10]. Toonimise tõhusus sõltub samas ka kasutatava mudeli arhitektuurist, kasutatavate mudeli parameetrite valikust ja treeningandmete kvaliteedist [13].

Oluline on märkida, et muusika andmete toonimine ei asenda originaalsete muusika andmete kasutamist. Samuti ei asenda toonimine masinõppe mudeli parameetrite ja algoritmide valimist. Parema tulemuse saavutamine eeldab tööd kõigil loetletud suundadel, et saavutada parimaid tulemusi muusikažanrite klassifitseerimise täpsuses.

Samuti on oluline märkida, et muusika žanrite definitsioonid ei ole üheselt kokku lepitud. Seetõttu ei ole ka teada, kuidas erinevad toonimised mõjutavad klassifitseerimise mudeli täpsust.

Käesolev magistritöö käsitleb muusika andmetike toonimise kasutamist muusika žanrite automaatse klassifitseerimise mudelite täpsuse parandamisel. Töö eesmärk on uurida, kuidas erinevad toonimise meetodid mõjuvad mudeli täpsusele ja erinevatele muusika žanritele.

## 2. Mõisted ja terminid

**Muusika andmeanalüütika** – MIR (ingl *music information retrieval*) on interdistsiplinaarne uurimisvaldkond, mis tegeleb muusikast informatsiooni leidmisega<sup>1</sup>.

**Andmete rikastamine** – (ingl *data augmentation*) on andmestiku suurendamine olemasolevate andmete modifitseerimise abil<sup>2</sup>.

**Muusika toonimine** – (ingl *music augmentation*) on autori poolt kasutusele võetud eestikeelne koondav termin muusikafailide erinevatele töötlustele.

**Tehisintellekt** - AI (ingl *artificial intelligence*) on tehnilise süsteemi võime hankida, töödelda ja rakendada teadmust sarnaselt inimintellektiga informatsiooni leidmisega<sup>3</sup>.

**Masinõpe** – ML (ingl *machine learning*) on protsess, mis kasutab andmetest või kogemustest õppimise võimaldamiseks arvutustehnilisi meetodeid informatsiooni leidmisega<sup>4</sup>.

**Süvaõpe** – DL (ingl *deep learning*) on masinõppe haru, mis imiteerib ülesannete lahendamisel inimajule iseloomulikku närvivõrkude struktuuri ehk tehisnärvivõrke informatsiooni leidmisega<sup>5</sup>.

---

<sup>1</sup> [https://en.wikipedia.org/wiki/Music\\_information\\_retrieval](https://en.wikipedia.org/wiki/Music_information_retrieval)

<sup>2</sup> <https://akit.cyber.ee/term/3930>

<sup>3</sup> <https://akit.cyber.ee/term/2183-tehisintellekt>

<sup>4</sup> <https://akit.cyber.ee/term/9968-masinope-automaatope>

<sup>5</sup> [https://en.wikipedia.org/wiki/Deep\\_learning](https://en.wikipedia.org/wiki/Deep_learning)

### 3. Taustainfo

#### 3.1 MIR - muusika andmeanalüütika

MIR (ingl *Music Information Retrieval* ehk MIR) on valdkond, mis kasutab arvutitehnoloogiat muusikaga seotud andmete töötlemiseks, analüüsimiseks ja esitamiseks. Eesti keeles on Tartu Ülikooli õppejõud Dr. Anna Aljanaki võtnud kasutusele termini „muusika andmeanalüütika“ [1]. MIR-i uurimisega tegelevad teadlased, insenerid ja muusikud üle maailma, kes püüavad mõista muusika olemust ja leida viise, kuidas seda arvutisüsteemide abil paremini tõlgendada ja kasutada.

MIR-i eesmärk on luua automatiseeritud tööriistu, mis suudavad muusikateoseid analüüsida ja neist erinevat tüüpi andmeid välja võtta. Näiteks võivad andmeteks olla heli signaal, rütm, meloodia, harmoonia, muusikaline struktuur, kuid ka oskus muusika instrumente tuvastada ja klassifitseerida muusika žanr. MIR võimaldab kasutajal seda teavet automaatselt eraldada, mis aitab omakorda tal paremini mõista muusika struktuuri ja sisu ning luua uusi muusikateoseid.

MIR on multidistsiplinaarne valdkond, mis ühendab erinevaid valdkondi, nagu muusikateooria, signaalitöötlus, statistika, masinõpe ja infotehnoloogia. MIR-i uurijad kasutavad erinevaid meetodeid ja tehnikaid, nagu digitaalsignaaltöötlus, masinõpe, andmeanalüüs, neuroteadus, et arendada välja tõhusad ja täpsemad MIR-algoritmid.

MIR-i rakendused on mitmekesised ja ulatuvad muusikaplatvormidest nagu Spotify ja Apple Music, kus kasutatakse algoritme muusika soovitamiseks ja leidmiseks, kuni interaktiivsete muusikaliste rakendusteni, mis võimaldavad inimestel luua oma muusikat ja eksperimenteerida erinevate helide ja rütmidega. MIR-i kasutatakse ka muusikateaduses ja muusikainstrumentide disainis, et paremini mõista muusika olemust ja aidata kaasa uute muusikaliste ideede ja innovatsioonide arengule.

MIR oluline valdkond, mis aitab meil paremini mõista muusika struktuuri ja sisu ning kasutada seda teavet mitmesugustes rakendustes, alates muusika loomisest ja esitamisest kuni muusika teaduslike uuringuteni.

#### 3.2 Muusika žanrite automaatne klassifitseerimine

Muusika žanrid on inimeste loodud kategoorilised sildid muusikateoste iseloomustamiseks. Muusika žanri iseloomustavad selle liikmete ühised omadused. Need omadused on tavaliselt seotud muusika instrumentatsiooni, rütmilise struktuuri ja harmoonilise sisuga. Žanrite hierarhiaid kasutatakse tavaliselt veebis saadaolevate suurte muusikakogude struktureerimiseks [2].

Muusika žanrite klassifitseerimine on oluline osa muusikatööstusest. Klassifitseerimine võimaldab eristada erinevaid muusikastiile ja aidata inimestel leida muusikat, mis vastab nende maitsele. Traditsiooniliselt on muusika žanrite klassifitseerimine põhinenud inimeste subjektiivsel hinnangul. See tähendab, et muusikatööstusel oli vaja kogenud inimesi, kes suudaksid muusikat kuulates eristada erinevaid žanreid, et märgendada muusikat.

Automaatne muusika žanri klassifitseerimine võib selles protsessis inimest aidata või asendada ning see oleks väärtuslik täiendus muusikateabe otsingusüsteemidele. Lisaks pakub automaatne muusika žanri klassifikatsioon raamistiku funktsioonide arendamiseks ja hindamiseks mis tahes tüüpi muusikasignaali sisupõhiseks analüüsiks [2].

Automatiseeritud muusika klassifitseerimiseks kasutatakse laialdaselt masinõpet, eriti süvaõpet (ingl *deep learning*). Süvaõpe põhineb masinaõppesüsteemidel, mis õpivad algoritmide abil ise ennustama ja kategoriseerima erinevaid muusikastiile. Süvaõpe toimib ka siis, kui muusika on esitatud kui helilaineid, spektraalanalüüsi või muusikaliste omaduste komplektina.

Sõltuvalt klassifitseerimise eesmärgist kasutatakse ka muusika tunnustel (ingl *feature engineering*) põhinevaid meetodeid, kus muusikateosest eraldatakse erinevad tunnusjooned nagu helitugevus, rütm, harmoonia, meloodia jne, mida seejärel kasutatakse masinõppe või süvaõppega muusikastiili määramiseks. See meetod ei ole nii täpne, kui kogu muusikaliste andmete komplekti kasutamine, kuid mudelite treenimine on oluliselt kiirem ja sõltuvalt konkreetsest seatud eesmärgist võib osutada oluliselt efektiivsemaks.

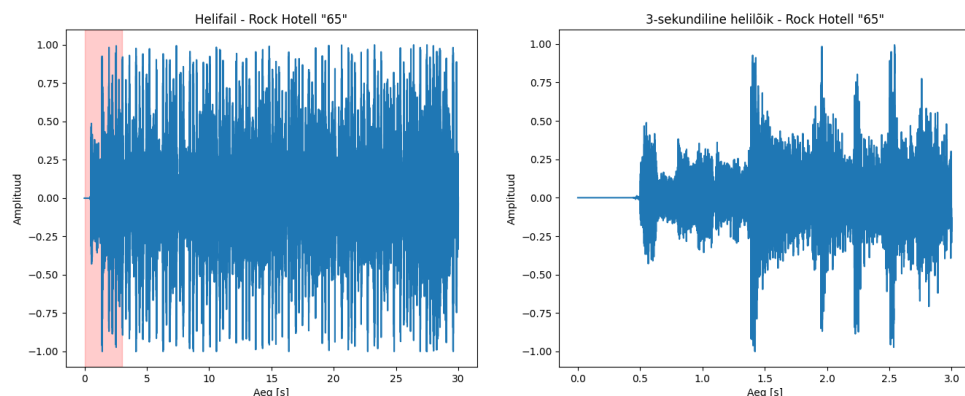
### 3.3 MFCC kasutamine

MFCC (ingl *Mel Frequency Cepstral Coefficients*) on helisignaali tunnused, mis saadakse helisignaali spektrist. MFCC töötati välja kõnekeele tämbri kirjeldamiseks ja neid tunnuseid kasutatakse ka muusika žanrite klassifitseerimisel [3].

MFCC tunnusteks on tavaliselt vektorite jada, mis esindavad helisignaali omadusi. Esialgsel helisignaalil võib olla tuhandeid või isegi miljoneid samplipunkte, kuid MFCC ekstraheerib nendest signaalidest olulisi omadusi ja esitab need vektorite jadana, kus igal vektoril on kindel arv omadusi.

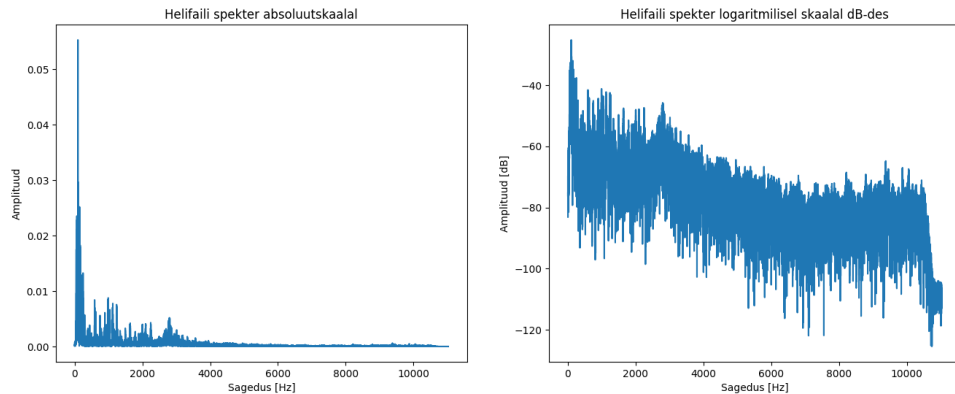
MFCC tunnuste leidmise protsess on järgmine [3]:

1. Helisignaali lõigatakse lühikesteks kaadriteks. Joonis 1 toob välja esimese 3 sekundi kadreerimise ja tulemuse ansambli „Rock Hotel“ loost „65“:



Joonis 1. Helifaili kadreerimine. *Rock Hotell* „65“.

2. Igale kaadritele rakendatakse signaali spektri leidmiseks Fourier' teisendust. Joonis 2 toob välja *Rock Hotel*-i loo „65“ 3 sekundilise kaadri sagedusspektreid absoluut- ja logaritmilisel skaalal. Info paremaks edastamiseks kasutatakse logaritmilist skaalat.



Joonis 2. Helifaili sagedusspektrid.

3. Mel skaala rakendamine. Inimese kuulmistaju ei ole lineaarne. Erinevates helistiketes olevate nootide erinevust tajuvad inimesed samasugustena, kuigi nootide sageduste vahed on erinevad. Näiteks nootide C4 ja D4 erinevus tundub samasugune, kui C5 ja D5 erinevus, kuid nootide sageduste vahe on erinev:

$$D4[293,66\text{Hz}] - C4[261,33\text{Hz}] = 32,03\text{Hz}$$

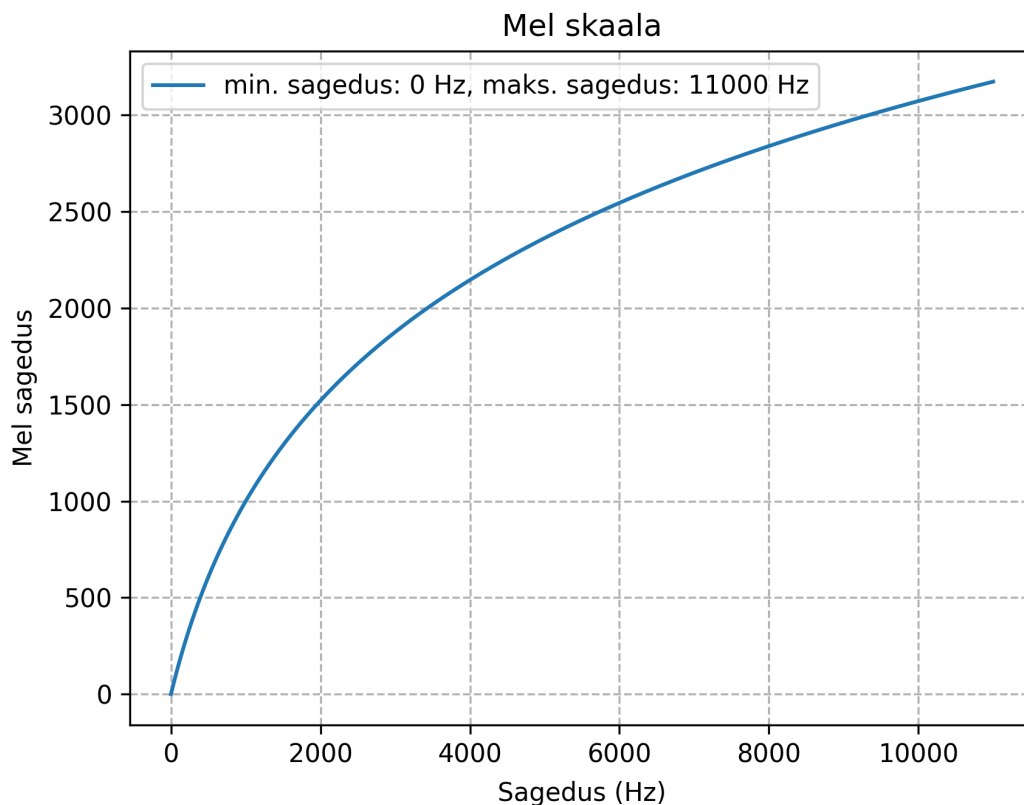
$$D5[587,33\text{Hz}] - C5[523,25\text{Hz}] = 64,08\text{Hz}$$

Inimese kuulmistaju imiteerimiseks kasutatakse Mel skaalat [4]. Sageduse  $f$  teisendamiseks Mel sageduseks  $m$  kasutatakse valemit [1]:

$$m = 2595 \log_{10} \left( 1 + \frac{1}{700} \right) \quad [1]$$

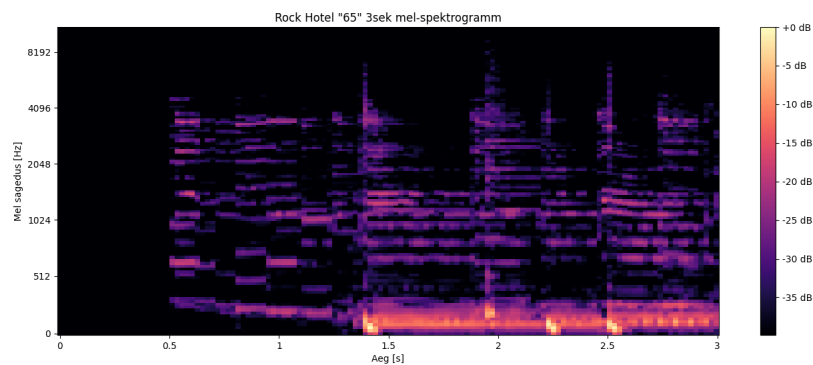
Joonisel 3 on esitatud Mel skaala vahemikus 0...11000Hz.





Joonis 3. Mel skaala.

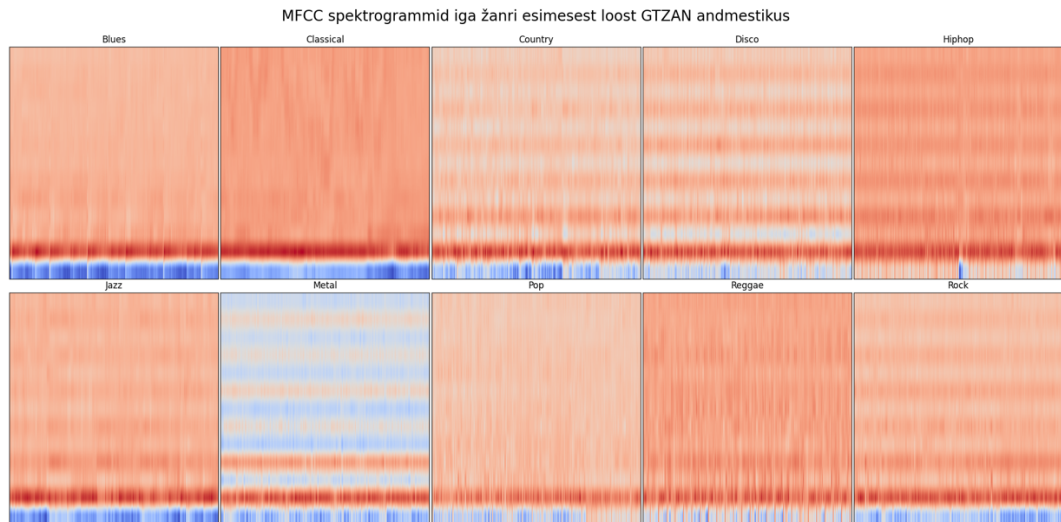
4. Arvutatakse Mel-spektrogramm. Selleks jagatakse sagedusvahemik Mel-filtritega lõikudeks. Mel-filtrite sagedusvahemik suureneb logaritmiliselt. Mel-filtrite arv *num\_mfcc* määrab spektrogrammi eraldusvõime. Joonisel 4 on näha *Rock Hotel*-i loo „65“ 3 esimese sekundi Mel-spektrogrammi.



Joonis 4. Mel-spektrogramm. Rock Hotel „65“.

5. MFCC tunnuste ehk MFCC koefitsientide arvutamine. Mel-spektrogramm sisaldab väga palju informatsiooni, sellel põhinevad mudelid vajavad palju arvutusressurssi. Seetõttu on mudeleid keeruline ja ajamahukas treenida. Tihti on otstarbekam rakendada Mel-spektrogrammile diskreetset koosinuse teisenduse DCT (ingl *Discrete Cosine Transform*) funktsiooni ja arvutada selle põhjal MFCC koefitsiendid [5].

Joonisel 4 on kujutatud *GTZAN* andmestiku eri žanride esimese loo MFCC spektrogrammid. Joonisel 5 on 10 juhusliku „*metal*“ žanri loo spektrogrammid. Visuaalne vaatlus toob välja sarnaseid mustreid samas žanris ja erinevusi eri žanrite spektrogrammide vahel. Spektrogrammil nähtavate horisontaalsete vöötide arvu määrab MFCC parameeter *num\_mfcc*.



Joonis 4. *GTZAN* andmestiku iga žanri esimese loo MFCC spektrogrammid.



Joonis 5. *GTZAN* andmestiku „*metal*“ žanri 10 juhusliku loo MFCC spektrogrammid.

MFCC koefitsiendid on olulised tunnused muusika žanrite klassifitseerimisel, sest võimaldavad kirjeldada muusikat väikeses dimensiooniruumis ja filtreerida soovimatuid helisid.

### 3.4 Muusika andmestikud. *GTZAN* andmestik

Muusika andmeanalüütikas on aluseks kasutatav andmestik. Üks parimaid organiseeritud muusika andmestike kogumikke internetis on Alexander Lerchi poolt koostatud ja hallatud kogumik [6], mis sisaldab üle 240 erineva muusika andmestiku tasuta allalaadimiseks kasutamiseks.

Teadusartiklites leiavad neist enim mainimist *GTZAN*, *Million Song Dataset – MSD*, *Latin Music Dataset – LMD*, *MagnaTagTune* ja *ISMIR*.

Magistritöö põhineb *GTZAN* andmestikul. Andmestiku autor on G.Tzanetakis, esmakordselt avaldati andmestik 2002 [2]. Andmestik sisaldab 1000 30 sekundit pikka .wav formaadis muusikaklippi. Esindatud on 10 žanri (märgendid: *blues*, *classical*, *country*, *disco*, *hip-hop*, *jazz*, *metal*, *pop*, *reggae*, *rock*), igas žanris 100 lugu. Kõikide klippide puhul on kasutatud diskreetimissagedust 22050Hz. Kõik heliklipid on monofoonilised. Helisignaali amplituud on kvanteeritud 16 bitilisena.

Andmestik on igale soovijale kättesaadav [7] ja on laialt leidnud kasutust paljudes muusika žanrite klassifitseerimise ülesannete sisendina. Masinõppe võistluste keskkonnas Kaggle [8] on *GTZAN*-i kasutatud enam kui 100 võistlustöö puhul [9].

### 3.5 Muusika andmestike rikastamine

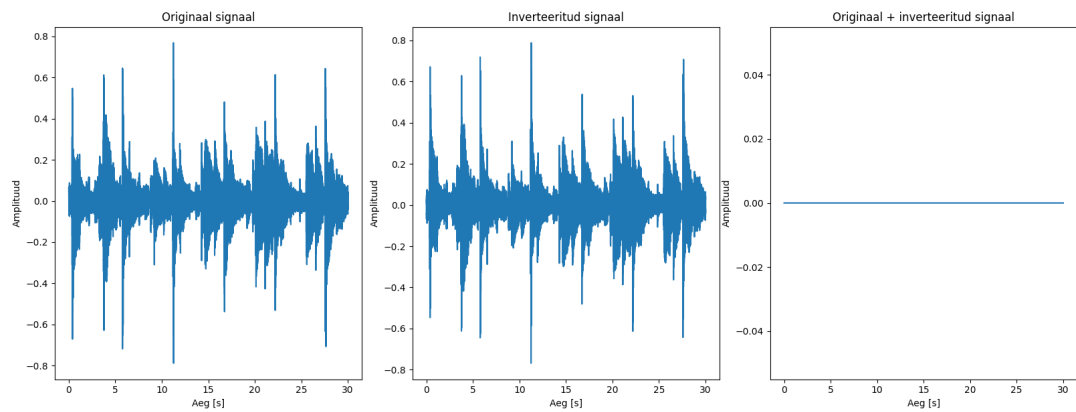
Muusika andmestike rikastamine (ingl *music data augmentation*) on tehnoloogia, mille all mõeldakse olemasolevate muusika andmestike suurendamist kasutades andmestikus olevate muusikafailide toonimist (ingl *music augmentation*) ja lisamist andmestikele kui uued failid. See meetod võimaldab andmestikke suurendada ilma uusi muusikapalasisid lisamata ja selleläbi suurendada andmestikku mahtu ja andmestikku kasutavate mudelite täpsust.

Kuigi muusika andmestike rikastamine võib anda muusikažanrite klassifitseerimisel kaasa tuua suurema täpsuse, on oluline meeles pidada, et see tehnoloogia võib mõne žanri puhul anda vastupidise tulemuse.

### 3.6 Muusikafailide toonimine

Muusikafailide toonimine on meetod, kus muusikafailidele rakendatakse mitmesuguseid deformatsioone, efekte jms. Levinumad toonimised on:

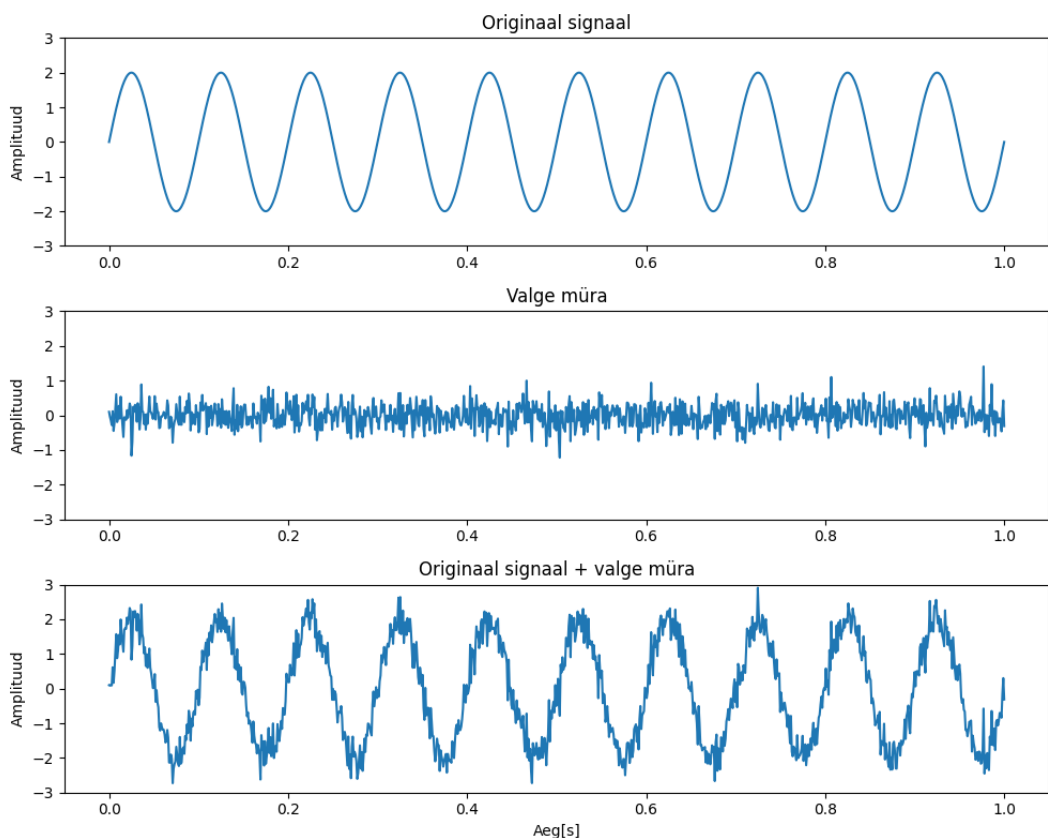
1. Helisignaali invertteerimine. Invertteerimist kasutatakse praktikas laialdaselt mürasummutavates kõrvaklappides välise müra kompenseerimiseks. Joonisel 6 esitab sisendsignaali, sisendsignaalist invertteeritud signaali ja nende summaarset signaali.



Joonis 6. Signaali inverteerimine ja summeerimine.

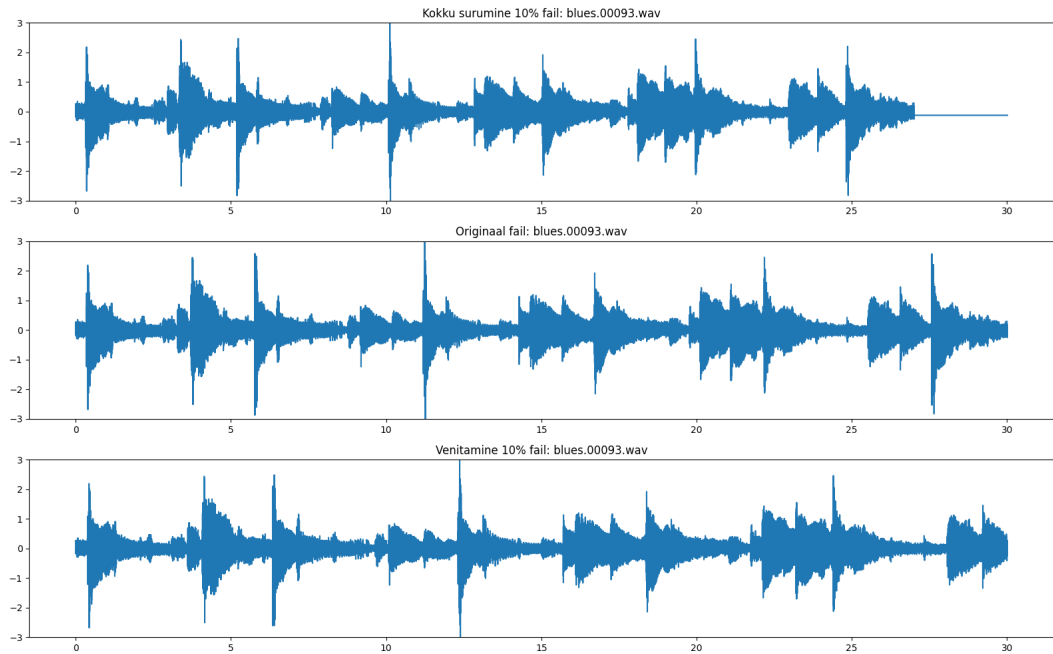
Helisignaali inverteerimine ei moonuta helisignaali. Inimese kõrv ei suuda eristada inverteeritud signaali originaalsignaalist.

2. Helile valge müra lisamine. Valge müra on juhuslik signaal, millel on erinevatel sagedustel võrdne intensiivsus, mistõttu on valgel müral konstantne spektraaltihedus. Joonisel 7 on kujutatud signaali, valget müra ja nende summat.



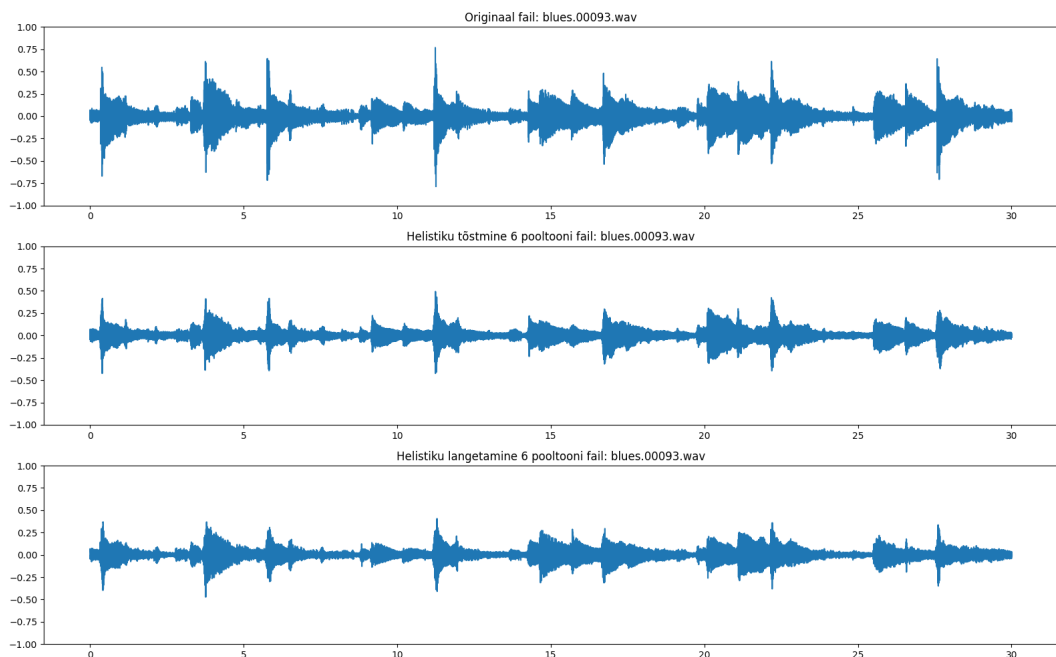
Joonis 7. Signaali ja valge müra summa.

3. Heli ajaline kokku surumine ehk „kiiruse tõstmine“ ja venitamine ehk „kiiruse vähendamine“ (ingl *time stretching*). Heli ajaline kokku surumine muudab heli kõrgemaks ja kiiremaks. Heli venitamine muudab heli madalamaks ja aeglasemaks. Joonis 8 toob välja 10% võrra kokku surutud helipildi, originaal helipildi, 10% võrra venitatud helipildi. Selgelt on näha helipildi liikumine ajas paremale.



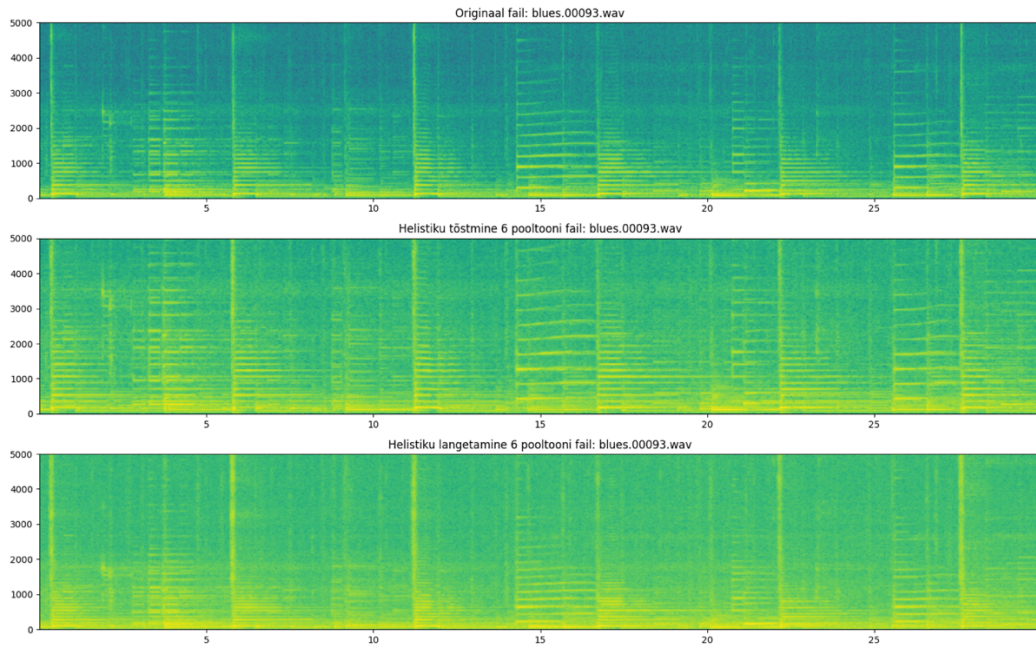
Joonis 8. Heli venitamine ja kokku surumine.

4. Helistiku tõstmine ja langetamine (ingl *pitch shift*) pooltoonide võrra. Erinevalt heli ajaliselt kokku surumisest või venitamisest jääb helistiku tõstmise või langetamise puhul jääb loo pikkus samaks. Joonisel 9 on kujutatud 30 sekundilist originaal-, 6 pooltooni tõstetud ja 6 pooltooni langetatud helina.



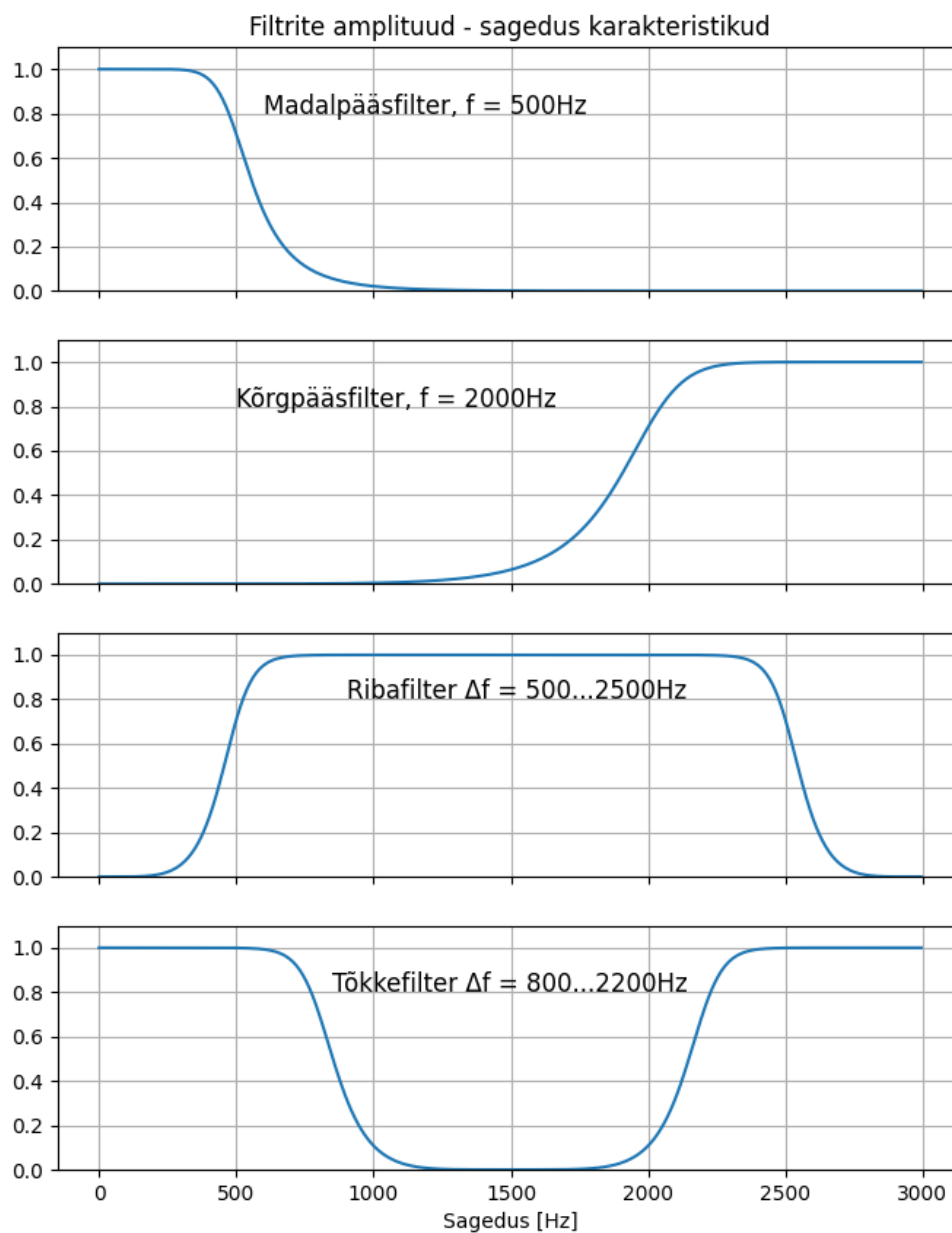
Joonis 9. Helistiku tõstmine ja langetamine. Helifailide võrdlus.

Helipiltide visuaalseks võrdlemiseks sobib paremini spektrogrammide võrdlus. Joonisel 10 on originaal helifaili spektrogramm ning tõstetud ja langetatud helistikuga helifailide spektrogrammid. Helistiku tõstmine suurendab võimsustihedust kõrgematel sagedustel ja langetamine madalamatel sagedustel.



Joonis 10. Helistiku tõstmine ja langetamine. Spektrogrammide võrdlus.

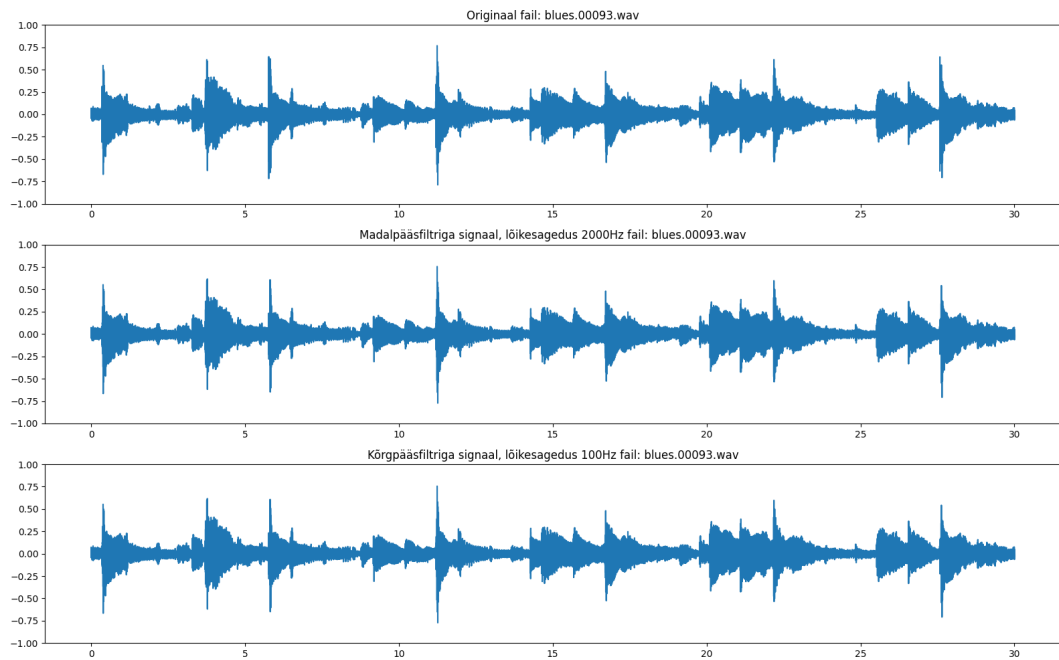
5. Helile mitmesuguste filtrite rakendamine (madalpääs-, kõrgpääs-, riba- ja tõkkefiltrid). Madalpääsfiltrid lasevad läbi sagedusi, mis on alla nende piirsageduse, ja nõrgendavad järk-järgult sagedusi, mis ületavad piirsagedust. Kõrgpääsfilter toimib vastupidiselt, laseb läbi kõrgeid sagedusi, mis on üle piirsageduse, ja nõrgendab järk-järgult piirsagedusest madalamaid sagedusi. Ribafilter laseb läbi sagedused oma kahe piirsageduse vahel, summutades samal ajal vahemikust väljapoole jäävaid sagedusi. Tõkkefilter nõrgendab sagedusi oma kahe piirsageduse vahel, lastes samal ajal läbi need, mis jäävad väljapoole piirsagedusi. Joonisel 11 on kujutatud erinevate filtritüüpide amplituud-sageduse karakteristikuid.



Joonis 11. Erinevat tüüpi filtrite amplituud-sagedus karakteristikud.

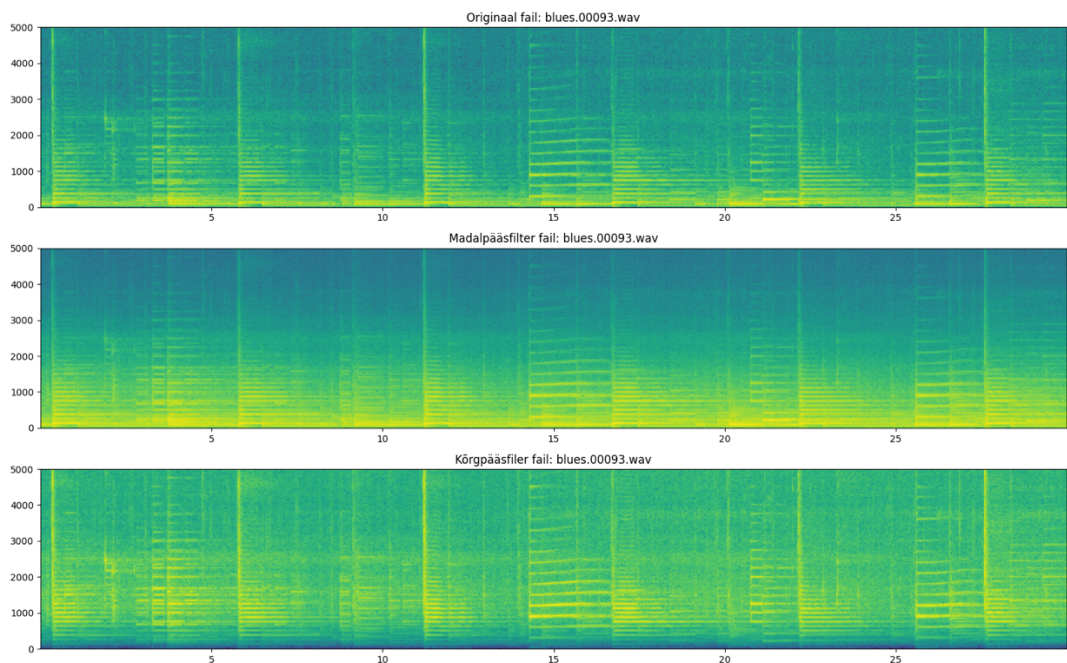
Filtrite kasutamist on helipildilt raske märgata. Näiteks joonisel 12 on kujutatud signaal ja tema väljund madalpääs- ja kõrgpääsfiltri läbimisel.





Joonis 12. Madal- ja kõrgpääsfiltrite mõju signaalile.

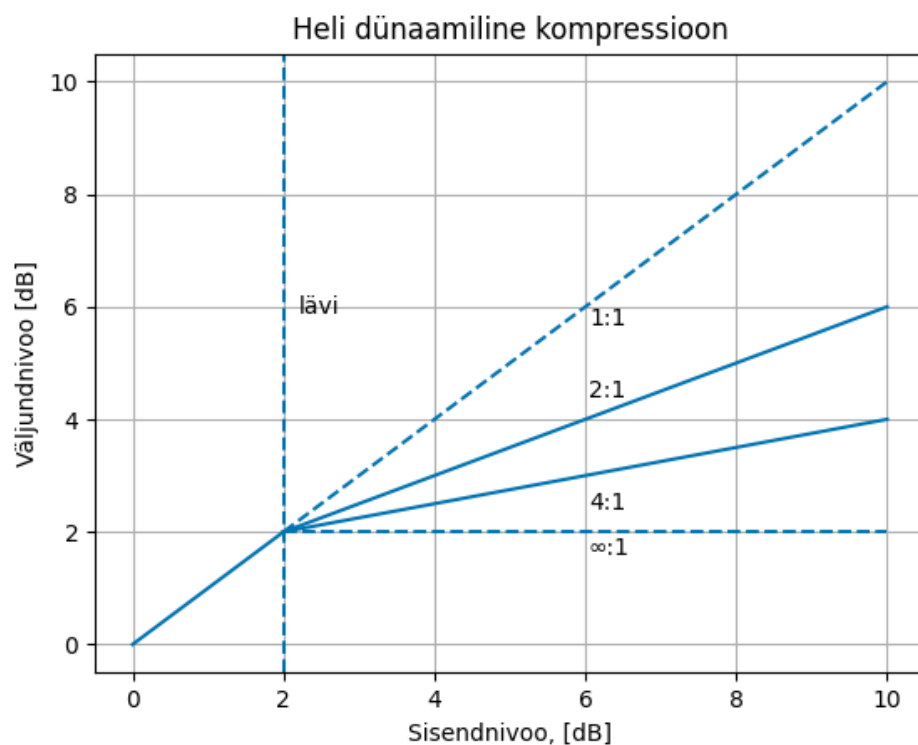
Filtrite mõju helisignaalile annab visuaalselt paremini edasi spektrogrammi kasutamine. Joonisel 13 on kujutatud signaali ja tema madalpääs- ja kõrgpääsfiltrite läbimisel tehtud spektrogrammid. Visuaalselt on näha madalpääsfiltri puhul signaali võimsuse langus kõrgematel sagedustel ja kõrgpääsfiltri puhul madalamatel sagedustel.



Joonis 13. Filtrite spektrogrammid.

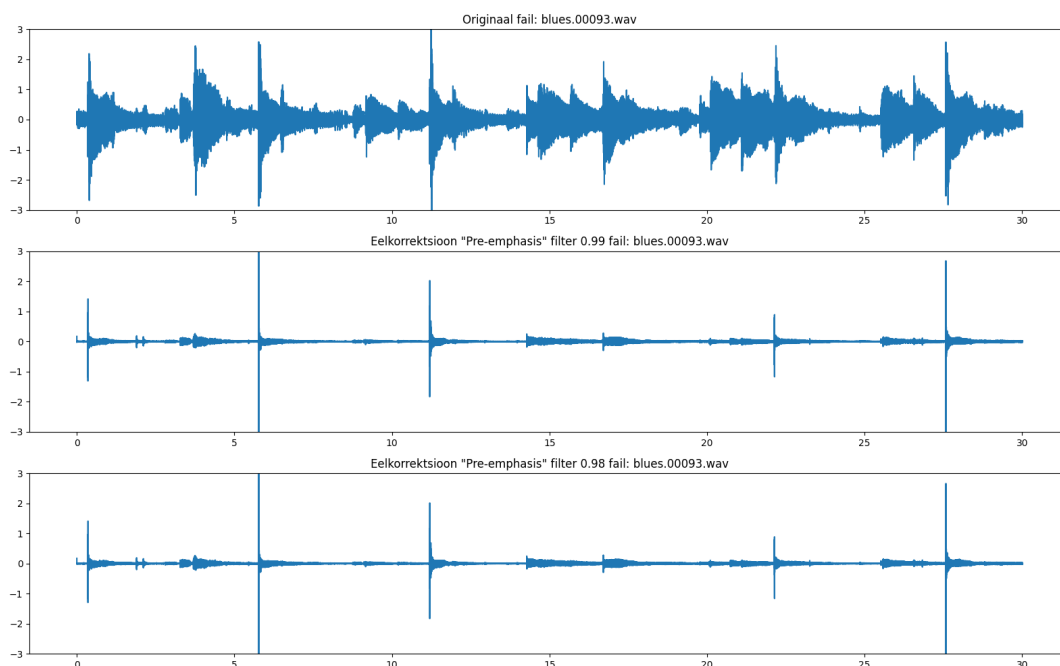


6. Heli dünaamiline kompressioon (ingl *dynamic range compression*) on helisignaali töötlemise toiming, mis vähendab valjude helide helitugevust või võimendab vaiksemaid helisid, vähendades või suurendades helisignaali dünaamilist ulatust. Kompressor vähendab helisignaali taset, kui selle amplituud ületab teatud läve. Tavaliselt seatakse lävi detsibellides. Kui signaal on alla läve, siis tihendamist ei toimu. Võimenduse vähendamise suurus määratakse suhtega. Näiteks suhe 4:1 tähendab, et kui sisendsignaali tase on 4 dB üle läve, väheneb väljundsignaali tase 1 dB-ni üle läve. Kõrgeimat suhet  $\infty:1$  nimetatakse piiravaks ja see tähistab, et iga läve ületav signaal vähendatakse lävetasemeni. Joonisel 14 on näidatud dünaamilise kompressiooni karakteristikat erinevate suhetega läve nivool 2 dB.



Joonis 14. Heli dünaamiline kompressioon.

7. Heli eelkorrektsioon (ingl *pre-emphasis*) on helisignaali töötlemine, mis rõhutab kõrgeid sagedusi ja vähendab madalamaid sagedusi. See on kasulik paljudes rakendustes, näiteks FM saatjates kompenseerimaks müra kõrgematel sagedustel või magnetofonides lindimüra vähendamiseks. Eelkorrektsioon võib aidata vähendada müra, parandada signaali selgust ja tugevdada kõrgemaid sagedusi. Eelkorrektsiooni kasutamisel näiteks lindistamisel tuleb lindistuse kuulamisel kasutada järelkorrektsiooni (ingl *de-emphasis*), et taastuks esialgne helipilt. Joonisel 15 on kujutatud eelkorrektsiooni mõju signaalile koefitsientidega 0,99 ja 0,98.



Joonis 15. Eelkorrektsiooni mõju originaal helile.

### 3.7 Varasemad uuringud

Tzanetakis ja Cook [2] uurisid muusikaliste žanrite klassifikatsiooni helisignaali põhjal. Autorid arendasid välja uue lähenemise, mis kasutab helisignaali ekstrahimiseks ja klassifitseerimiseks mitmesuguseid helianalüüsi meetodeid. Nende eesmärk oli automatiseerida žanri tuvastamine ja eristamine, mis on oluline muusika teabe haldamise, soovitusüsteemide ja digitaalse arhiveerimise jaoks. Autorite pakutud meetod on võimeline tuvastama ja eristama muusikalisi žanre, saavutades keskmise täpsuse umbes 61%. Uurimus aitas kaasa muusika informaatika valdkonna arengule, pakkudes täiustatud tehnikaid ja algoritme muusikaliste helisignaali analüüsimiseks ja žanri määramiseks.

Aguiar, Costa ja Silla [10] uurisid muusika andmete rikastamise meetodeid muusika žanri klassifitseerimisel kasutades spektrogramme. Signaalist genereeriti spektrogramm kõrgusega 256 pikslit ja spektrogrammi laius sõltus loo pikkusest. Spektrogramm lõigati 16 piksli laiusteks ribadeks ja ribad sisestati CNN (ingl *convolutional neural network*) põhisesse muusika klassifitseerimise mudelisse. Originaalandmestikuga (artiklis kasutatakse ingliskeelset terminit *Baseline*) saavutati mudeli täpsuseks 86,33%. Muusika toonimise meetoditest kasutati müra lisamist (mudeli täpsus vähenes 83,78%-ni), heli nivooga (ingl *loudness*) manipuleerimist (täpsus paranes minimaalselt, 86,44%-ni) helistiku tõstmist ja langetamist (ingl *pitch shift*) ühe (täpsus paranes 88,33%-ni) ja kahe pooltooni võrra (täpsus paranes 89,45%-ni) ja heli kokku surumist (täpsus paranes 86,78%-ni). Parimad tulemused saavutati *Latin Music Database*-i kasutades mistõttu pole tulemused otseselt võrreldavad *GTZAN* andmestikuga.

Cheng, Chang ja Kuo [11] kasutasid oma muusika žanrite klassifitseerimise töös *GTZAN* andmestikku ja Mel-spektrogramme. Mudel põhines 5 kihiga CNN-mudelil, kus igal kihil kasutati *dropout*-i tasemega 0,5. Parimaks täpsuseks saavutati 83,30%. Huvipakkuv fakt nende tööst on mudeli treenimisele kulunud aeg. Mudeli treenimisel kasutati NVIDIA RTX 2080Ti GPU serverit mäluga 64GB, mida võib tänasel päeval (mai 2023) võrrelda kahe väga

hea PC-põhise mänguarvuti võimekusega. Mudeli treenimisele (kokku 2180 kordust) kulus 13,1 tundi.

Liang ja Gu [12] kasutasid oma töös muusika žanrite klassifitseerimiseks eeltreenitud *Musicnn* mudeleid. *Musicnn* sisaldab mitut eeltreenitud mudelit nagu *MTT\_musicnn* (treenitud *MagnaTagTune* andmestikul), *MSD\_musicnn* ja *MSD\_musicnn\_big* (mõlemad treenitud *Million Song Dataset*-il). Eeltreenitud *Musicnn* mudelite kasutamine võimaldab oluliselt kiiremini tulemusteni jõuda ja tänu nende treenimisele oluliselt suurematel andmestikel (ca 1000 korda suuremad kui *GTZAN*) võimaldavad ka oluliselt parema täpsuse. Oma töös saavutasid nad mudeli keskmise täpsuse 88,36%. *Musicnn*-i mudelitega treenimisel kasutatakse *GTZAN*-iga võrreldes veidi erinevaid žanrite klassifitseerimist. Lisanduvad *rap*, *folk*, *electronic* ja *r&b*, kuid puuduvad *disco*, *hip-hop* ja *reggae*, kokku on mudelis 11 žanri. Autorid toovad välja žanrite üheselt mõistetavate definitsioonide puudumise probleemi ja selle mõju klassifitseerimise mudeli täpsusele.

Schindler, Lidy ja Rauber [13] võrdlesid oma töös erinevaid muusika žanrite klassifitseerimise mudeleid (*Shallow* ja süvaõppe) erinevatel andmestikel (*GTZAN*, *ISMIR Genre*, *Latin Music Database*, *Million Song Dataset*). Toonimistel kasutati heli kokku surumist (koefitsiendid 1,2 ja 1,5), heli venitamist (koefitsiendid 0,2 ja 0,5), helistiku tõstmist ja langetamist (mõlemal juhul 2 ja 5 pooltooni võtta). *GTZAN* andmestikuga saavutati süvaõppe mudelit kasutades originaalandmestikuga täpsuseks 78,6% ja koos toonitud andmestikuga 81,8%.

## 4. Metoodika

Autor teostas uurimuse kasutades programmeerimiskeelt Python. Peamisteks tööriistadeks olid Pythoni teegid Pandas [14], Numpy [15], Librosa [16] ja PyTorch [17]. Uurimuses kasutatud metoodika on jagatud järgmisteks mooduliteks:

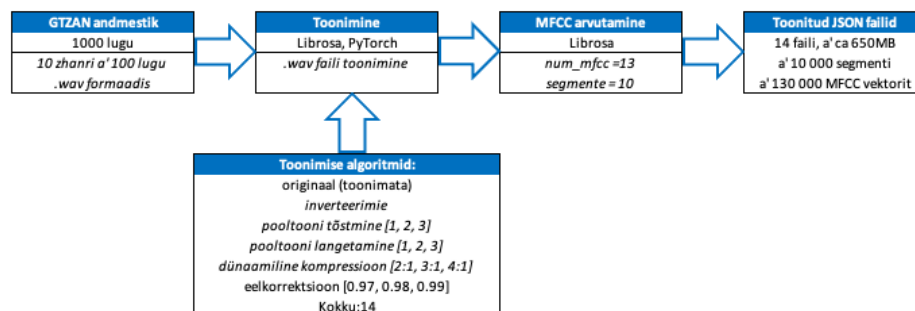
1. Töö andmestikuga. *GTZAN* andmestiku erinevate muusika toonimise meetoditega rikastamine optimaalsete MFCC parameetrite ja helifaili segmentide arvuga;
2. Muusika klassifitseerimise mudeli optimeerimine ja parameetrite valik *GTZAN* originaalandmestikul;
3. Erinevate toonimiste mõjude võrdlemine. Mudelite treenimine kombineeritult originaal- ja erinevate toonitud andmetega andmestikel.

### 4.1 Töö andmestikuga

Käesolevas töös kasutas autor järgmisi muusika toonimise meetodeid:

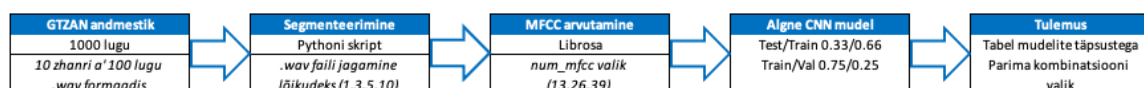
1. Helisignaali inverteerimine;
2. Helisignaali helistiku tõstmine 3-s versioonis: 1-, 2- ja 3-pooltooni võrra;
3. Helisignaali helistiku langetamine 3-s versioonis: 1-, 2- ja 3-pooltooni võrra;
4. Helisignaali eelkorrektsioon 3-s versioonis: koefitsientidega 0.99, 0.98 ja 0.97;
5. Helisignaali dünaamiline kompressioon 3-s versioonis: suhtarvud 2:1, 3:1 ja 4:1 lävendil -10dB.

Helisignaali toonimisel kasutas autor Librosa ja PyTorch teeki. Toonitud helifailid (1000 .wav formaadis faili igaüks 30 sekundit pikk) jagati 10-ks kolm sekundit pikaks segmentiks. Igale segmentile arvutati Librosa teeki kasutades MFCC koefitsiendid ja tulemus koondati iga toonimise kohta ühte JSON faili, mis olid klassifitseerimise mudeli üheks sisendiks. Joonis 16 kirjeldab toonimise ja MFCC koefitsientide arvutamise etappe.



Joonis 16. Helifailide toonimine ja MFCC koefitsientide arvutamine.

Sobivaimate MFCC filtrite arvu ( $num\_mfcc=13/26/39$ ) ja helifaili segmenteerimiste jaguri ( $segmente=1/3/5/10$ ) leidmiseks teostati algse mudeli treenimisel originaalandmestikuga kõikide  $num\_mfcc$  ja segmentide arvu kombinatsioonidega. Joonisel 17 on esitatud valiku protsess.



Joonis 17. MFCC filtrite arvu  $num\_mfcc$  ja helifailide segmentide arvu leidmine.

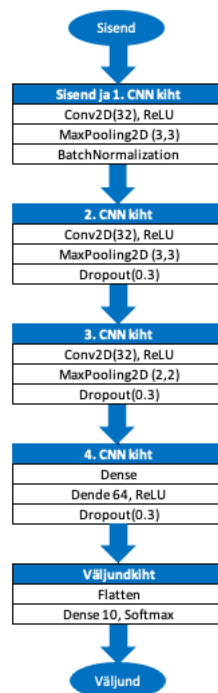
Parimaid tulemusi andnud MFCC filtrite arvu ( $num\_mfcc=13$ ) ja segmentide arvuga ( $segmente=10$ ) teostati MFCC koefitsientide arvutused kõikide toonimistega (kokku 13) ja originaalandmestikuga. Tulemuseks oli 14 JSON faili.

Algseks mudeliks kasutas autor Valerio Velardo veebikoolitusel [18] koostatud CNN-il põhinevat muusika klassifitseerimise mudelit. Mudeli täpsus oli 69%, mis on piisav nimetatud parameetrite valikuks.

## 4.2 Mudeli optimeerimine ja treenimine GTZAN originaalandmestikul

Toonimiste efekti hindamiseks tuleb kombineeritud originaal- ja toonitud andmetikega testimisel kasutada ainult originaalandmestikust moodustatud testandmeid. Algselt kasutatud Valerio Velardo mudel pole ideaalne. GTZAN andmestikul on MFCC tunnuste põhistel žanrite klassifitseerimistel saavutatud paremaid mudeli täpsusi kui 69%. Paganakis, Kotropoulus ja Arce [19] toovad oma töös välja erinevatest teadusartiklitest koondatud tabeli mudelite täpsustega alates 61,0% kuni 82,5%. Varasemates uurimustes on kasutatud ka erinevaid mudeli arhitektuure.

Autor kasutas brutaalse rünnaku meetodit leidmaks parim kombinatsioon CNN mudeli arhitektuurist ja parameetritest. Leitud mudeli täpsuseks originaalandmetega treenides kujunes 78,88%, mis oli 14,2% parem tulemus võrreldes aluseks olnud mudeliga. Autori mudeli arhitektuur on esitatud joonisel 18.

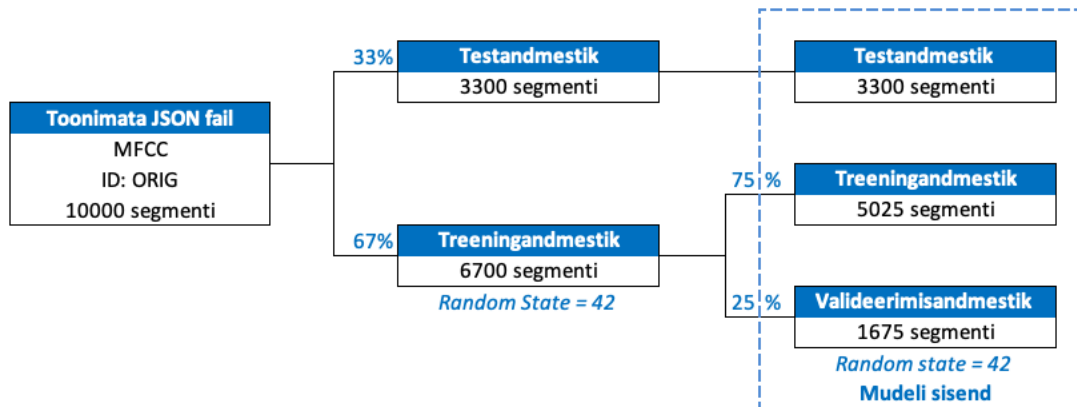


Joonisel 18. Autori koostatud CNN mudel.

## 4.3 Erinevate toonimiste mõjude võrdlemine

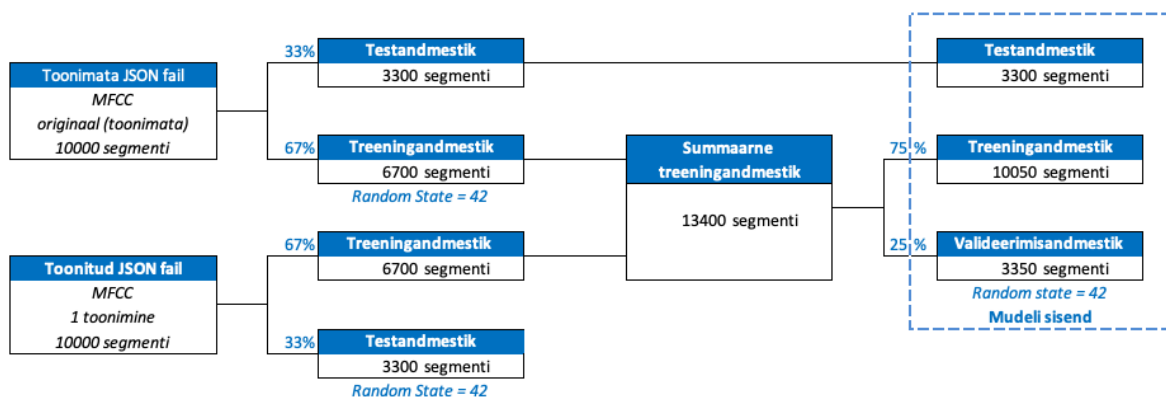
Erinevate toonimiste efekti hindamiseks võrreldakse toonitud mudeli täpsust originaalse andmestikuga treenitud etalon mudeliga. Etalon mudel treenitakse ainult

originaalandmestikku kasutades. Joonis 19 kirjeldab originaalandmestikust etalon mudeli sisendi moodustamist.



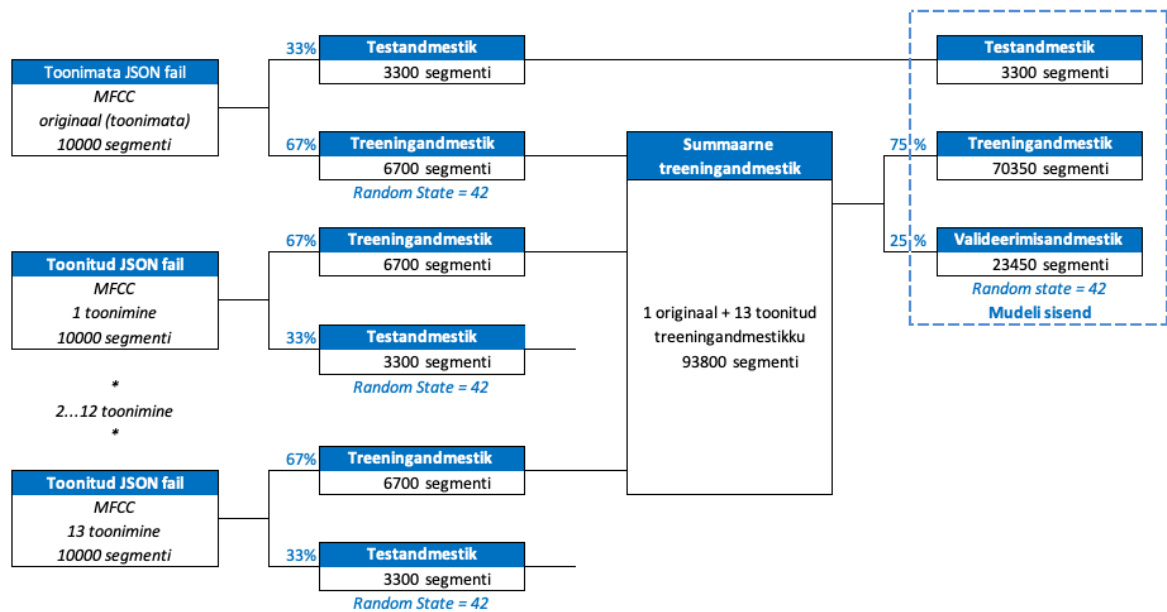
Joonis 19. Etalon mudeli sisendandmestik.

Erinevate toonimiste võrdluseks etalon mudeliga treeniti toonitud mudelit kombineeritud andmestikuga, mis koosnes toonimata ja toonitud andmestikest. On oluline, et testandmestik oleks sama mis etalon mudelil. Võrreldavate andmestike saamiseks kasutati kõikidel jagamistel sama *Random State* väärtust, mis tagab test- ja treeningandmestikeks jagamise alati sama juhuslikkusega. Toonitud andmestiku jagamisel tekkinud testandmestikku ei kasutatud. Joonisel 20 näeme kombineeritud andmestiku loomist.



Joonis 20. Kombineeritud andmestike jagamine mudeli test-, treening ja valideerimisandmestikeks.

Analoogselt kombineeritud andmestikuga koostati ka mudeli kõikide toonimistega korraka treenimise andmestik. Testandmestik võeti ainult toonimata andmestikust, treeningandmestik koostati kõikide andmestike treeningandmestikest. Kõikide toonimiste kombineeritud andmestiku koostamist selgitab joonis 21.



Joonis 21. Toonimata ja kõikide toonitud andmestike jagamine mudeli test-, treening ja valideerimisandmestikeks.

Kõikide mudelite treenimiste väljundiks olid:

- JSON fail, mis sisaldas informatsiooni mudeli treenimise ajatemplist, mudeli toonimistest, kasutatud parameetritest ja mudeli täpsusest;
- Segadusmaatriks tekstifailina;
- Segadusmaatriks graafilise .png failina;
- Mudeli täpsuse ja kadude graafik graafilise .png failina.

Tulemuste võrdluseks ja analüüsiks kasutas autor Pythonis kirjutatud programmi, mis kogus kõikide mudelite JSON failidest andmed ühte Microsoft Exceli formaadis tabelisse.

## 5. Tulemused

### 5.1 Erinevate toonimiste meetoditega saavutatud mudelite täpsused

Käesolevas töös kasutati järgmisi muusika toonimise meetodeid:

1. Helisignaali inverteerimine;
2. Helisignaali helistiku tõstmine 3-s versioonis: 1-, 2- ja 3-pooltooni võrra;
3. Helisignaali helistiku langetamine 3-s versioonis: 1-, 2- ja 3-pooltooni võrra;
4. Helisignaali eelkorrektsioon 3-s versioonis: koefitsientidega 0.99, 0.98 ja 0.97;
5. Helisignaali dünaamiline kompressioon 3-s versioonis: suhtarvud 2:1, 3:1 ja 4:1 lävendil -10dB.

Kõikide meetodite tulemused on esitatud žanrite kaupa käesoleva töö lisas:

- Täpsuste tabelina;
- Segadusmaatriksina;
- Täpsuse ja kao graafikuna mudeli treenimiste kordustel.

Tabel 1 kirjeldab erinevate toonimiste mõju žanrite klassifitseerimiste täpsustele.

Tabel 1. Toonimiste mõju žanritele klassifitseerimiste täpsusele.

Žanr	Dünaamiline kompressioon	Helistiku langetamine	Helistiku tõstmine	Eelkompressioon	Heli inverteerimine	Kõik toonimised koos ühes andmestikus	Etalon mudel
pop	5,9%	4,8%	5,4%	10,1%	4,6%	7,5%	82,1%
metal	-4,8%	0,4%	-3,3%	-4,9%	1,6%	-2,4%	91,0%
disco	6,2%	6,4%	9,0%	10,7%	3,5%	9,8%	72,8%
blues	0,4%	-0,9%	2,1%	-5,7%	-2,0%	-1,3%	90,3%
reggae	-0,4%	0,5%	5,0%	-0,8%	-13,4%	-2,5%	72,5%
classical	1,8%	3,8%	2,9%	4,8%	1,7%	3,2%	94,3%
rock	8,8%	15,8%	5,6%	19,8%	23,1%	29,1%	50,3%
hiphop	-2,8%	-0,7%	-7,3%	-3,8%	10,2%	-1,1%	81,7%
country	20,5%	18,4%	29,4%	22,1%	32,2%	39,8%	61,6%
jazz	-0,5%	-2,6%	-2,3%	-7,3%	-0,6%	-2,9%	92,2%
Mudeli täpsus	81,0%	81,7%	81,8%	81,3%	82,6%	83,5%	78,9%

Töö tulemusel selgus, et võrreldes etalon andmestikuga treenitud mudeliga:

- kõik toonimistega treenitud mudelid parandasid ennustuse keskmist täpsust üle kõikide žanrite enam, kui 2% võrra;
- kõik toonimised andsid parema tulemuse žanrites *pop*, *disco*, *classical*, *rock* ja *country*;
- kõik toonimised andsid halvema tulemuse žanris *jazz*, enam kui pooltel toonimistel saavutati halvem tulemus žanrites *metal*, *blues*, *reggae* hip-hop;



- ei esinenud ühtegi toonimist, mis andnuks parema tulemusega kõikides žanrites;
- kõikide toonimiste puhul ilmnes halvem mudeli täpsus vähemalt kolmes žanris.

## 5.2 Tulemuste analüüs

Tulemus, et kõik toonimised andsid keskmiselt parema tulemuse, ei anna alust väitmaks, et muusika toonimine on alati töökindel meetod mudeli täpsuse parandamisel. Väga palju sõltub seatud eesmärgist, valitud mudelist ja kasutatavast andmestikust. Autori töö käigus tehtud varasemad MFCC põhised mudelid andsid tulemusi, kus mõningad toonimised andsid etaloniga võrreldes halvema keskmise täpsuse, kuid samas oli täpsus erinevate žanrite vahel ühtlasem. Samuti annab MFCC tunnuste kasutamine erinevaid tulemusi võrreldes Mel-spektrogrammidel põhinevate mudelite kasutamisega. Autor kasutas oma töö alguses faasis MFCC tunnuste asemel Mel-spektrogramme ja väga suurel andmestikul eelnevalt treenitud *Musicnn* [28] mudeleid, mille puhul toonimiste tulemused erinesid MFCC-l põhinevatel mudelitel saadutega. *Musicnn* mudelite keskmine täpsus jäi autoril alla 60%, mis polnud toonimiste võrdluste tegemiseks piisav täpsus ja autor jätkas edaspidi vaid MFCC-l põhinevate mudelitega.

Andmestikes kasutatud märgendid žanrite kohta on andmestiku koostaja subjektiivne arvamus. Erinevad muusika žanrite definitsioonid pole üheselt määratletud. Mitmed muusika žanrid on tekkinud teisest žanrist väikeste modifikatsioonidega. Näiteks *rock* tekkis 1950-ndatel USA lõunaosariikides *blues*-ist. Mängiti *blues* -i lugu, samade nootidega, kuid kiirema tempoga ja seda nimetati *rock*-iks. Margus Kappel esitab ansambli Rock Hotel ajalugu käsitlevas filmis [21] efektse näite, kui vähe erinevad omavahel *blues* ja *rock*. Muusikud pole kokku leppinud, millisest tempost alates lõpeb *blues* ja algab *rock*. Samuti on hajusad piirid žanrite *rock*, *country*, *pop* vahel. Näiteks žanr *pop* tuleneb terminist populaarne muusika, mille definitsioon muutub ka ajas.

Subjektiivsus märgendite määramisel võis olla peamine põhjus, miks etalon mudelil oli žanril *rock* täpsus vaid 50,3% ja *country*-l 61,6%. Inimene arvab omamoodi, mudel leiab aga erinevuse teistel alustel.

Žanrid *classic* ja *jazz* käitusid mudelite toonimisel erinevalt. Mõlema žanri täpsus etalon andmestikul oli väga kõrge, vastavalt 94,3% ja 92,2%. *Classic* žanri tulemus paranes kõikide toonimistega, *jazz*-i täpsus vastupidi, vähenes kõikide toonimistega. Ühe põhjusena võib ära märkida, et mõlemad erinevad oma stiililt ülejäänud žanritest, heli kõlapilt sama žanri sees erineb väga oluliselt. Andmestiku autor Tzanetakis on teinud oma töös [2] ettepaneku jagada need stiilid omakorda alamklassidesse. Žanr *classic* tuleks tema arvates jagada järgmiselt: *choir*, *orchestra*, *piano* ja *string quartet* ja žanr *jazz* jagada järgmiselt: *bigband*, *cool*, *fusion*, *piano*, *quartet* ja *swing*. Vastavate märgendite puudumisel andmestikus ei olnud autoril võimalik uurida, kuidas selline jagamine oleks muutnud mudeli täpsust erinevate toonimistega.

## 6. Kokkuvõte

Käesoleva töö tulemusena selgus, et kasutades muusika toonimist *GTZAN* muusika andmestikul, MFCC tunnustel põhineva muusika žanrite klassifitseerimise mudeli täpsus paranes keskmiselt üle kõikide žanrite, kuid ei leidunud ühtset muusika toonimise meetodit, mille rakendamine tagaks alati kõikide žanrite puhul parema mudeli täpsuse.

Autor toob oma töö tulemuste analüüsis välja mitmeid asjaolusid, mida tuleb silmas pidada muusika toonimiste kasutamisel žanrite klassifitseerimiste mudelite täpsuste parandamiseks.

## 7. Viidatud kirjandus

- [1] Mait Lindpere, „Loomuliku keele töötlemise algoritmide kasutamine muusikasarnasuse leidmisel,“ Magistritöö, Tartu Ülikool, 2022, [Online]. [https://comserv.cs.ut.ee/ati\\_thesis/datasheet.php?id=75074](https://comserv.cs.ut.ee/ati_thesis/datasheet.php?id=75074). (08.11.2022).
- [2] George Tzanetakis, Perry Cook, „Musical genre classification of audio signals,“ IEEE Transactions on speech and audio processing, Volume 10, Issue 5, (2002, July) 293-302. [Online]. <https://ieeexplore.ieee.org/document/1021072>. (01.02.2023).
- [3] MFCC. Mel-Frequency Cepstrum Coefficients, [Online]. <https://learn.flucoma.org/reference/mfcc/>. (15.03.2023).
- [4] Librosa.mel\_frequencies, [Online]. [https://librosa.org/doc/main/generated/librosa.mel\\_frequencies.html](https://librosa.org/doc/main/generated/librosa.mel_frequencies.html). (15.03.2023).
- [5] Speech Processing for Machine Learning: Filter banks, Mel-Frequency Cepstral Coefficients (MFCCs) and What's In-Between, [Online]. <https://haythamfayek.com/2016/04/21/speech-processing-for-machine-learning.html>. (20.03.2023.)
- [6] Datasets. Datasets for audio content analysis and music information retrieval, [Online]. <https://www.audiocontentanalysis.org/datasets.html>. (10.10.2022).
- [7] Gtzan. Github repository, [Online]. <https://github.com/tensorflow/datasets/blob/master/docs/catalog/gtzan.md>. (10.10.2022).
- [8] Kaggle [Online]. <https://www.kaggle.com/>. (22.01.2023).
- [9] GTZAN Dataset – Music Genre Classification [Online]. <https://www.kaggle.com/datasets/andradaolteanu/gtzan-dataset-music-genre-classification>. (22.01.2023).
- [10] Rafael L. Aguiar, Yandre M. G. Costa, Carlos N. Silla Jr., Exploring Data Augmentation to Improve Music Genre Classification with ConvNets,“ 2018 International Joint Conference on Neural Networks (IJCNN), (2018, July), [Online]. <https://ieeexplore.ieee.org/document/8489166>. (17.04.2023).
- [11] Yu-Huei Cheng, Pang-Ching Chang, Che-Nan Kuo „Convolutional Neural Networks Approach for Music Genre Classification,“ International Symposium on Computer, Consumer and Control (IS3C), (2020, November), [Online]. <https://ieeexplore.ieee.org/document/9394067>. (17.04.2023).
- [12] Beici Liang, Minwei Gu, „Music Genre Classification Using Transfer Learning,“ 2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), (2020, August), [Online]. <https://ieeexplore.ieee.org/document/9175547>. (17.4.2023)
- [13] Alexander Schindler, Thomas Lidy, Andreas Rauber, „Comparing Shallow versus Deep Neural Network Architectures for Automatic Music Genre Classification,“ Conference: 9th Forum Media Technology (FMT2016), (2016, November), [Online]. [https://www.researchgate.net/publication/313895490\\_Comparing\\_Shallow\\_versus\\_Deep\\_Neural\\_Network\\_Architectures\\_for\\_Automatic\\_Music\\_Genre\\_Classification](https://www.researchgate.net/publication/313895490_Comparing_Shallow_versus_Deep_Neural_Network_Architectures_for_Automatic_Music_Genre_Classification). (9.04.2023).
- [14] Pandas, [Online]. <https://pandas.pydata.org/>. (12.02.2023).
- [15] NumPy, [Online]. <https://numpy.org/>. (12.02.2023).

- [16] Librosa. Audio and music processing in Python, [Online]. <https://librosa.org/>. (12.03.2023).
- [17] PyTorch, [Online]. <https://pytorch.org/>. (12.02.2023).
- [18] Musikalkemist. Github repository, [Online]. <https://github.com/musikalkemist/DeepLearningForAudioWithPython>. (11.03.2023).
- [19] Yannis Paganakis, Constantine Kotropoulos, Gonzalo R. Arce, „Music genre classification via sparse representations of auditory temporal modulations,“ 2009 17th European Signal Processing Conference, (2009, August), [Online]. <https://ieeexplore.ieee.org/document/7077429>. (14.04.2023).
- [20] Musicnn. Github repository, [Online]. <https://github.com/jordipons/musicnn>. (10.10.2022).
- [21] Harry Kõrvits, „Rock Hotel. Kuidas läks sul see mäng,“ dokfilm (2023). ETV 27.01.2023, 20:00.

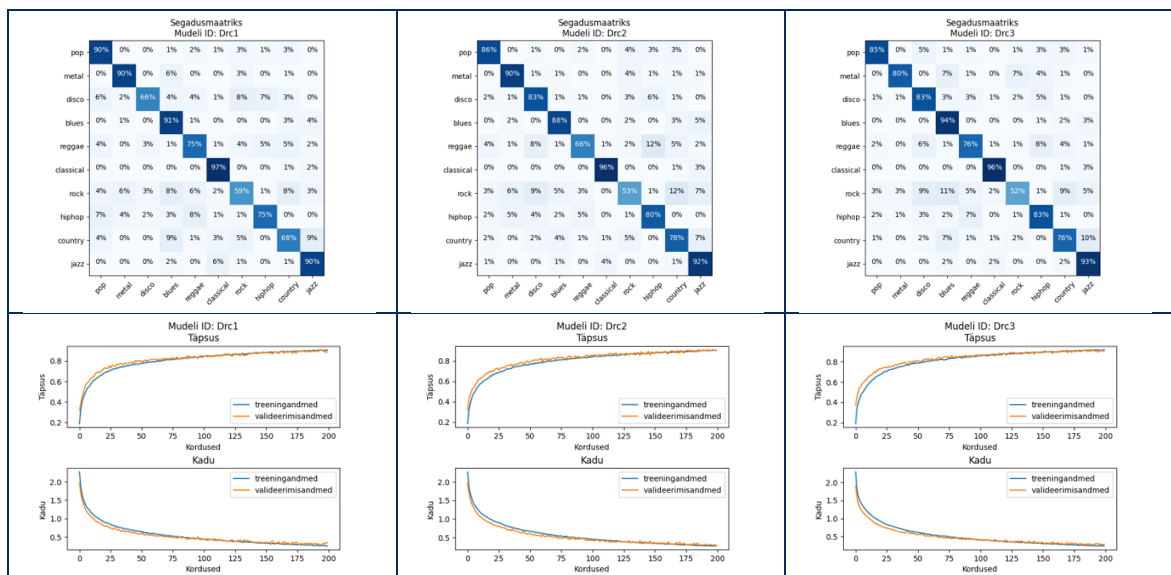
# Lisad

## I. Tulemuste tabelid, graafikud ja segadusmaatriksid

### 1. Heli d naamiline kompressioon

D naamiline kompressioon [Drc]

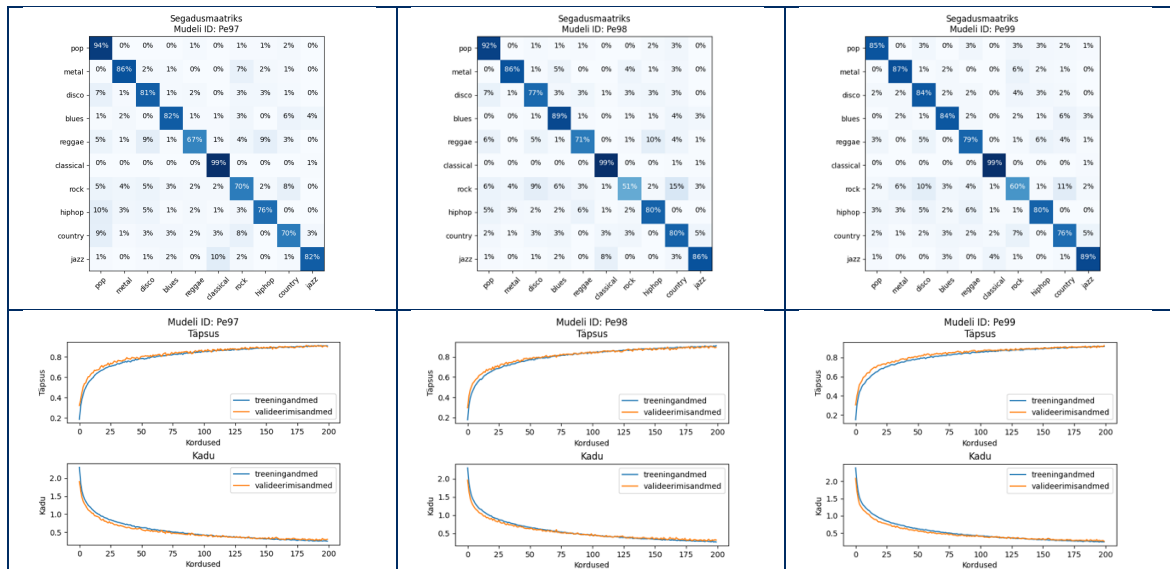
�anr	Drc1	Drc2	Drc3	Drc keskmine	Etalon	Vahe	Muutus
pop	89,6%	85,8%	85,3%	86,9%	82,1%	4,8%	5,9%
metal	89,8%	90,1%	79,9%	86,6%	91,0%	-4,4%	-4,8%
disco	65,5%	83,2%	83,2%	77,3%	72,8%	4,5%	6,2%
blues	90,9%	87,5%	93,6%	90,7%	90,3%	0,4%	0,4%
reggae	75,3%	65,8%	75,6%	72,2%	72,5%	-0,3%	-0,4%
classical	96,8%	95,6%	95,6%	96,0%	94,3%	1,7%	1,8%
rock	58,8%	53,5%	51,9%	54,7%	50,3%	4,4%	8,8%
hiphop	74,9%	80,4%	83,0%	79,4%	81,7%	-2,3%	-2,8%
country	68,2%	78,1%	76,3%	74,2%	61,6%	12,6%	20,5%
jazz	89,5%	92,2%	93,4%	91,7%	92,2%	-0,5%	-0,5%
	<b>79,9%</b>	<b>81,2%</b>	<b>81,8%</b>	<b>81,0%</b>	<b>78,9%</b>		



## 2. Heli eelkompressioon

### Eelkompressioon

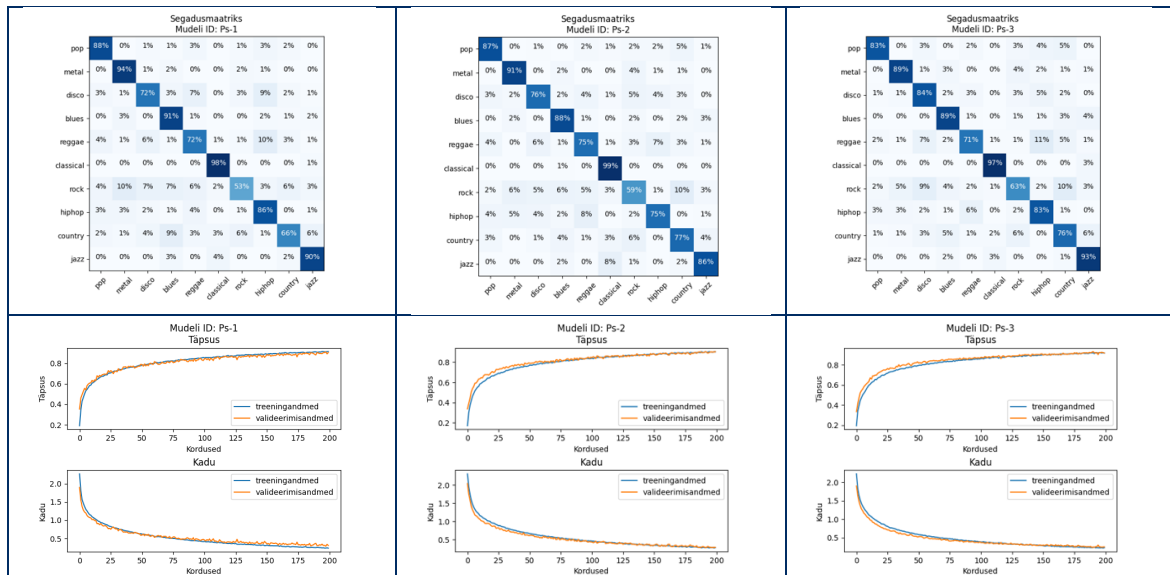
Žanr	Pe97	Pe98	Pe99	Pe keskmine	Etalon	Vahe	Muutus
pop	94,2%	91,9%	85,0%	90,4%	82,1%	8,3%	10,1%
metal	86,2%	85,9%	87,4%	86,5%	91,0%	-4,5%	-4,9%
disco	81,3%	76,9%	83,5%	80,6%	72,8%	7,8%	10,7%
blues	82,1%	89,4%	83,9%	85,1%	90,3%	-5,2%	-5,7%
reggae	66,7%	70,6%	78,6%	71,9%	72,5%	-0,6%	-0,8%
classical	99,0%	98,7%	98,7%	98,8%	94,3%	4,6%	4,8%
rock	69,8%	50,9%	60,1%	60,3%	50,3%	10,0%	19,8%
hiphop	76,2%	79,7%	79,7%	78,6%	81,7%	-3,1%	-3,8%
country	70,3%	79,6%	75,7%	75,2%	61,6%	13,6%	22,1%
jazz	81,7%	85,6%	89,2%	85,5%	92,2%	-6,7%	-7,3%
	<b>80,8%</b>	<b>80,9%</b>	<b>82,2%</b>	<b>81,3%</b>	<b>78,9%</b>		



### 3. Helistiku langetamine

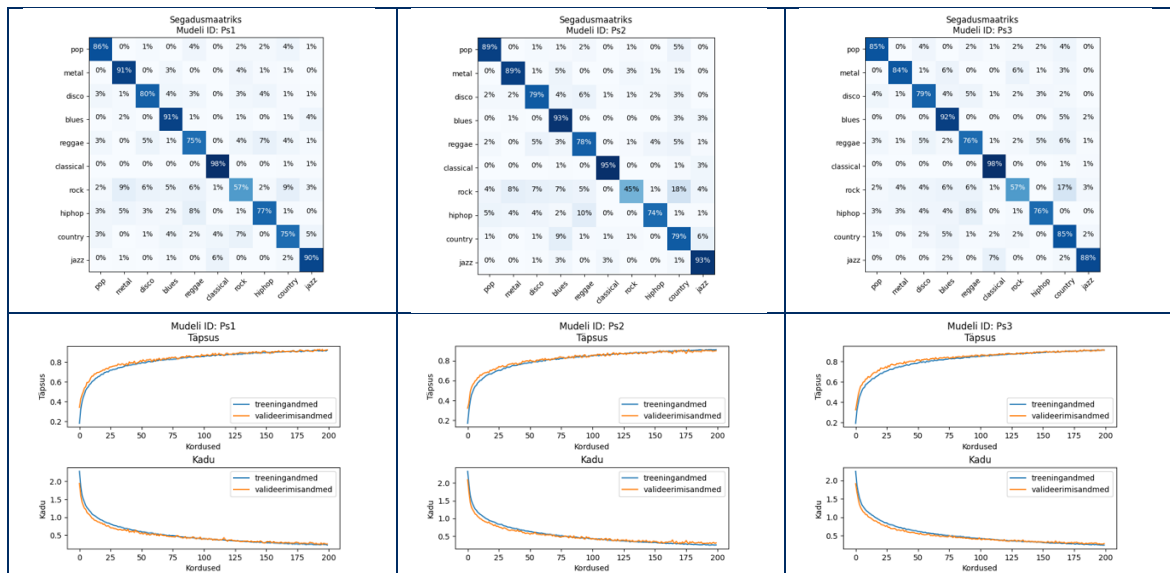
Helistiku langetamine [Ps-]

Žanr	Ps-1	Ps-2	Ps-3	Ps- keskmine	Etalon	Vahe	Muutus
pop	88,2%	87,0%	82,9%	86,0%	82,1%	3,9%	4,8%
metal	94,0%	91,0%	89,2%	91,4%	91,0%	0,4%	0,4%
disco	72,5%	76,3%	83,5%	77,4%	72,8%	4,6%	6,4%
blues	91,2%	88,4%	88,8%	89,5%	90,3%	-0,8%	-0,9%
reggae	72,2%	75,3%	71,1%	72,9%	72,5%	0,4%	0,5%
classical	98,4%	98,7%	96,5%	97,9%	94,3%	3,6%	3,8%
rock	52,8%	59,4%	62,6%	58,3%	50,3%	8,0%	15,8%
hiphop	85,5%	75,2%	82,6%	81,1%	81,7%	-0,5%	-0,7%
country	65,8%	77,2%	75,7%	72,9%	61,6%	11,3%	18,4%
jazz	90,4%	86,2%	92,8%	89,8%	92,2%	-2,4%	-2,6%
	<b>81,1%</b>	<b>81,5%</b>	<b>82,6%</b>	<b>81,7%</b>	<b>78,9%</b>		



#### 4. Helistiku tõstmine

Helistiku tõstmine[Ps+]							
Žanr	Ps1	Ps2	Ps3	Ps+ keskmine	Etalon	Vahe	Muutus
pop	86,1%	88,7%	84,7%	86,5%	82,1%	4,4%	5,4%
metal	91,3%	89,2%	83,5%	88,0%	91,0%	-3,0%	-3,3%
disco	79,7%	79,1%	79,1%	79,3%	72,8%	6,5%	9,0%
blues	91,2%	93,3%	92,1%	92,2%	90,3%	1,9%	2,1%
reggae	74,7%	77,8%	75,8%	76,1%	72,5%	3,6%	5,0%
classical	98,1%	94,9%	98,1%	97,0%	94,3%	2,8%	2,9%
rock	56,9%	45,3%	57,2%	53,1%	50,3%	2,8%	5,6%
hiphop	77,5%	73,6%	75,9%	75,7%	81,7%	-6,0%	-7,3%
country	75,4%	79,0%	84,7%	79,7%	61,6%	18,1%	29,4%
jazz	89,5%	92,5%	88,3%	90,1%	92,2%	-2,1%	-2,3%
	<b>82,1%</b>	<b>81,3%</b>	<b>81,9%</b>	<b>81,8%</b>	<b>78,9%</b>		





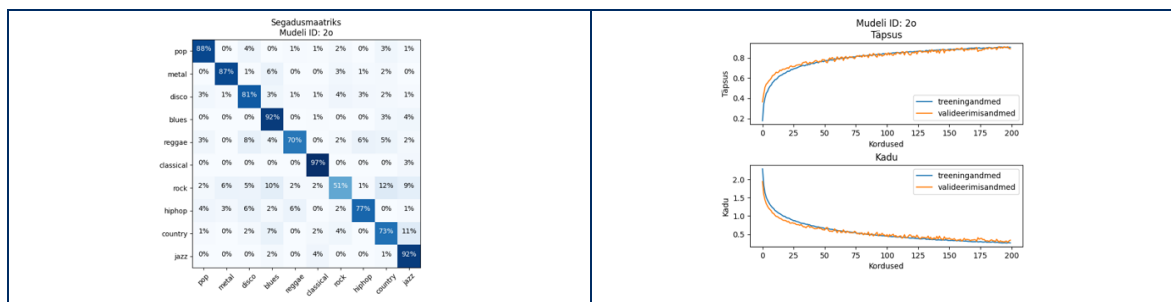
## 5. Heli inverteerimine

Heli inverteerimine [Pi]				
Žanr	Pi	Etalon	Vahe	Muutus
pop	85,8%	82,1%	3,8%	4,6%
metal	92,5%	91,0%	1,5%	1,6%
disco	75,3%	72,8%	2,5%	3,5%
blues	88,4%	90,3%	-1,8%	-2,0%
reggae	62,8%	72,5%	-9,7%	-13,4%
classical	95,9%	94,3%	1,6%	1,7%
rock	61,9%	50,3%	11,6%	23,1%
hiphop	90,0%	81,7%	8,4%	10,2%
country	81,4%	61,6%	19,8%	32,2%
jazz	91,6%	92,2%	-0,6%	-0,6%
	<b>82,6%</b>	<b>78,9%</b>		

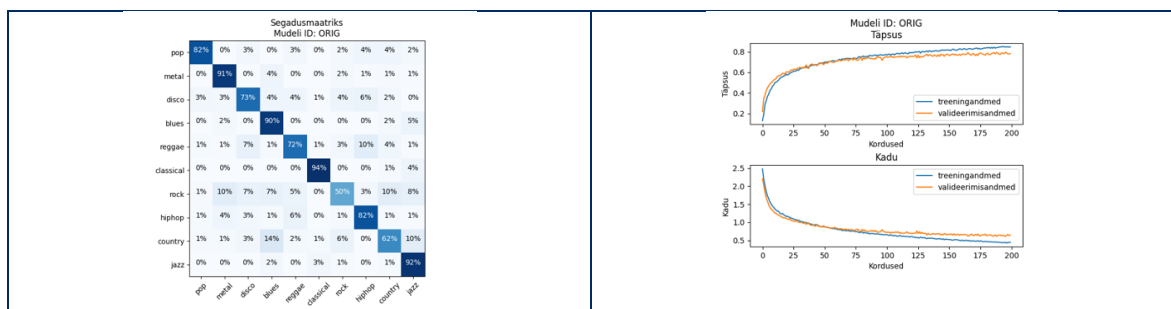


## 6. Kõik toonimised koos

Kõik toonimised korraga				
Žanr	Kõik koos	Etalon	Vahe	Muutus
pop	88,2%	82,1%	6,2%	7,5%
metal	88,8%	91,0%	-2,2%	-2,4%
disco	79,9%	72,8%	7,2%	9,8%
blues	89,1%	90,3%	-1,2%	-1,3%
reggae	70,7%	72,5%	-1,8%	-2,5%
classical	97,3%	94,3%	3,0%	3,2%
rock	64,9%	50,3%	14,6%	29,1%
hiphop	80,8%	81,7%	-0,9%	-1,1%
country	86,1%	61,6%	24,5%	39,8%
jazz	89,6%	92,2%	-2,7%	-2,9%
	<b>83,5%</b>	<b>78,9%</b>		



## 7. Etalon model

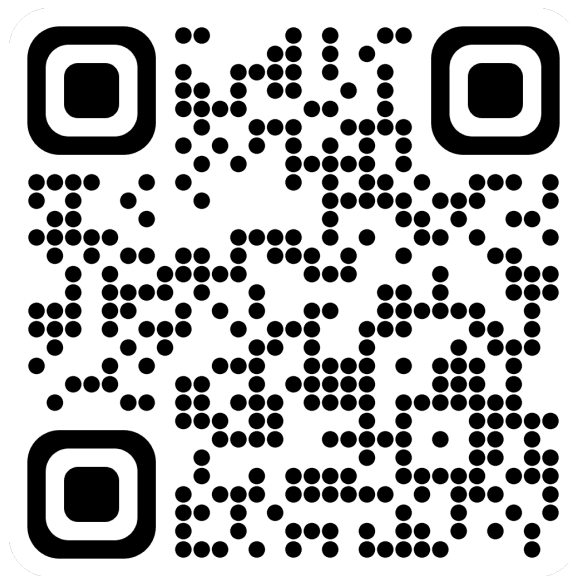


## II. Kasutatud tarkvara lähtekoodid

Kõik autori poolt loodud ja töös kasutatud programmide koodid on avalikult kättesaadavad autori Github-i kontol:

<https://github.com/RaivoKasepuu/magister>

Autori Github konto QR kood:



### III. Litsents

#### **Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks**

Mina, Raivo Kasepuu,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose „Muusika toonimise kasutamine muusika žanrite klassifitseerimise mudelites“ mille juhendaja on Anna Aljanaki reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.
2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

*Raivo Kasepuu*  
**08.05.2023**