

Julius Alamäki

# HARMONISEN JA PERKUSSIIVISEN ÄÄNEN EROTTELUMENETELMÄ JA SEN SOVELLUSKOHTEET

Informaatioteknologian ja viestinnän tiedekunta  
Kandidaattitutkielma  
Kesäkuu 2022

# TIIVISTELMÄ

Julius Alamäki: Harmonisen ja perkussiivisen äänen erottelumenetelmä ja sen sovelluskohteet  
Kandidaattitutkielma  
Tampereen yliopisto  
Tietojenkäsittelytieteiden tutkinto-ohjelma  
Kesäkuu 2022

---

Harmonisen ja perkussiivisen äänen erottelumenetelmällä jokin musikaaliseksi signaaliksi mielletty pyritään jakamaan sen harmoniseen ja perkussiiviseen komponenttiin. Harmonisen signaalin tulisi sisältää vain melodiset soivat äänet, kuten laulut, huilut, torvet ja viulut. Perkussiivisen signaalin taas tulisi sisältää erilaiset perkussiot kuten rummut. Tutkielmassa tarkastellaan mihin HPSS-menetelmä perustuu ja minkälaisien ongelmien ratkaisuun sitä on käytetty. Tavoitteena on myös selvittää, onko HPSS-menetelmää mahdollista käyttää reaaliaikaisesti äänen käsittelyyn ja osana niin sanotun ”sampler-instrumentin” toiminnallisuutta.

HPSS-menetelmän kannalta oleellisia työkaluja ovat erilaiset Fourier-muunnokset ja spektogrammi. Fourier-muunnoksen avulla äänidata saadaan siirrettyä aikatasolta spektritasolle, joka on tarpeellista, jos halutaan saada tietoa äänen sisältämistä taajuuksista. Spektogrammi voidaan ajatella olevan äänen visuaalinen esitys. Spektogrammin avulla voidaan tarkastella millaisia perkussiivisiä ja harmonisia ominaisuuksia äänisignaalilla on ja soveltaa monia kuvankäsittelystä tuttuja algoritmeja kuten mediaanisuodatinta äänidataan.

Fourier-muunnoksesta käydään läpi sen diskreetin esityksen periaatteet ja spektogrammi käydään läpi yleisellä tasolla. HPSS-menetelmästä esitetään yksi mahdollinen toteutus, jossa signaalin harmoninen ja perkussiivinen signaali erotellaan toisistaan mediaanisuodattimien ja binäärimaskien avulla.

Tutkimus ei suoraan osoita, että reaaliaikainen äänenkäsittely on mahdollista, mutta viittaa siihen. Mediaanisuodattaminen ja STFT-muunnoksen (Short-Time Fourier Transform) käyttö harmonisen ja perkussiivisen signaalin erottelussa, vaikuttaisi olevan yksi tehokkaimmista vaihtoehdoista. Vaikka reaaliaikainen äänen käsittely ei olisikaan mahdollista, harmonisen ja perkussiivisen äänen erottelumenetelmää voisi käyttää osana ”sampler-instrumentin” toiminnallisuutta.

Avainsanat: Harmonisen ja perkussiivisen äänen erottelumenetelmä, HPSS, signaalin käsittely, sampler-instrumentti

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck –ohjelmalla.

## Sisällysluettelo

<b>1</b>	<b>Johdanto .....</b>	<b>1</b>
<b>2</b>	<b>Tutkimusmenetelmä .....</b>	<b>2</b>
<b>3</b>	<b>HPSS-menetelmästä .....</b>	<b>3</b>
3.1	Fourier-muunnos	4
3.2	Spektogrammi	6
3.3	Harmonisen ja perkussiivisen signaalin erottelu	7
<b>4</b>	<b>Sovellukset .....</b>	<b>9</b>
4.1	MIR-sovellukset	9
4.1.1	Sointutunnistus	9
4.1.2	Melodian eristys	10
4.1.3	Äänenväriin muokkaus ja korvaaminen	10
4.2	Muita sovelluksia	11
4.2.1	Sydän- ja hengityssäänen mittaaminen	11
4.2.2	Sisäkorvaistutteen käyttäjien musiikinkuuntelukokemuksen parantaminen	12
4.2.3	Puheensävyjen tunnistaminen	12
<b>5</b>	<b>Pohdinta .....</b>	<b>13</b>
<b>6</b>	<b>Yhteenveto .....</b>	<b>14</b>
	<b>Lähdeluettelo .....</b>	<b>16</b>

## LYHENTEET JA MERKINNÄT

BSS	Blind Source Separation
DFT	Discrete Fourier Transform
FFT	Fast Fourier Transform
HMM	Hidden Markov Model
HPSS	Harmonic-percussive Sound/Source Separation
ISTFT	Inverse Short-Time Fourier Transform
MIR	Music Information Retrieval
NMF	Non-Negative Matrix Factorization
SNMF	Shifted Non-Negative Matrix Factorization
STFT	Short-Time Fourier Transform
VAD	Voice Activity Detection

## 1 Johdanto

Musikaaliset signaalit koostuvat usein hyvin erilaisista äänilähteistä. Yksi tapa erotella äänilähteitä, on jakaa ne harmonisiin ja perkussiivisiin äänilähteisiin. Harmonisia äänilähteitä ovat esimerkiksi erilaiset kielisoittimet kuten piano, kitara ja viulu. Perkussiivisilla äänilähteillä tarkoitetaan usein lyömäsoittimia kuten perinteisiä rumpuja ja niiden eri osia kuten virveliä ja bassorumpua. (Tachibana ym., 2012) Näillä erityyppisillä signaaleilla on hyvin erityyppisiä ominaisuuksia ja niitä voidaankin erotella toisistaan menetelmällä, jota kutsutaan ”harmonic/percussive sound separation” -menetelmäksi ja siitä käytetään lyhennettä (HPSS). Menetelmää hyödyntäen voidaan myös pyrkiä tunnistamaan musikaalisessa signaalissa esiintyviä sointuja tai rytmisiä jaksoja tai taajuuskorjaamaan signaalin harmonisia ja perkussiivisiä elementtejä erikseen. (Ras & Wieczorkowska, 2010) HPSS-menetelmää on myös kokeiltu soveltaa lääketieteellisessä tarkoituksessa, erottamaan sydän- ja hengitystä toisistaan (Torii ym., 2019).

Tutkielmassa pyritään selvittämään, mihin HPSS-menetelmä perustuu ja millaisia työkaluja sen toteuttamiseksi tarvitaan. Lisäksi tutkitaan, että voidaanko HPSS-menetelmää käyttää reaaliaikaisen äänisignaalin käsittelyssä. Tämän tutkielman tuloksia ja sen tekemisen aikana opittuja asioita on tarkoitus hyödyntää tuotekehityksessä uuden ”sampler-instrumentin” suunnittelussa ja kehityksessä.

Sampler-instrumentin pääasiallinen tavoite on jäljitellä digitaalisesti, jotain oikeaa soitinta esimerkiksi pianoa, rumpuja tai viulua. Sampler-instrumentin toiminnallisuutta voidaan verrata perinteisiin syntetisaattoreihin. Syntetisaattori luo usein ääntä tuottamalla erilaisia, usein yksinkertaisia, jaksollisia aaltoja. Tällaisia aaltoja voi olla siniaalto, kanttiaalto tai saha-aalto. Johonkin matemaattiseen kaavaan perustuen syntetisaattori siis tuottaa ääntä niin sanotusti tyhjästä. Tämän vuoksi syntetisaattorilla on hyvin vaikea tuottaa ääntä, joka kuulostaisi joltain oikealta instrumentilta, koska soitettujen instrumenttien tuottama ääni ja sen satunnaisuus on hyvin vaikeaa esittää matemaattisena kaavana. Tähän ongelmaan ratkaisuksi on keksitty sampler-instrumentit.

Sampler-instrumenttiin syötetään matemaattisten kaavojen sijaan, vaikka pianon yksittäisten koskettimien äänet. Nämä äänet äänitetään ensin erillisiksi äänitiedostoiksi, jonka jälkeen ne järjestellään sampler-instrumenttiin niin, että ne vastaavat oikeita koskettimien paikkoja. Sampler-instrumentin valmistelussa voidaan äänittää myös yksittäisestä pianokoskettimesta eri versioita ja eri voimakkuudella soitettuja äänitteitä, jotta instrumentilla saadaan paremmin jäljitelyä oikean pianon dynaamisuutta ja satunnaisuutta. Tällä tavoin sampler-instrumentista voidaan saada ”inhimillisemmän” ja ”aidomman” kuuloinen. Nykyään löytyy jo paljon ”sampler-instrumentteja”, jotka imitoivat kokonaisia sinfoniaorkestereita. Näissä instrumenteissa pianon sijaan on

äänitetty erikseen eri orkesterin soitinryhmiä, joita voidaan jälkikäteen soittaa ”sampler-instrumentilla”. Elokuvamusiikissa käytetään jo hyvin yleisesti tällaisia ”sampler-instrumentteja” oikean orkesterin sijaan, sillä se on huomattavasti halvempaa, kuin oikeasti säveltää orkesterille ja äänittää sen soittoa. Tarkka kuuntelija voi kuitenkin kuulla, jos tällaiseen ratkaisuun on päädytty. ”Sampler-instrumentilla” tehty orkesterimusiikki kuulostaa usein liian puhtaalta ja liian tarkalta ollakseen oikeaa ihmisten soittoa. Tämä johtuu siitä, että sampler-instrumentteihin ei kyetä äänittämään satoja eriversioita jokaisesta erillisestä ja eri lailla soitetusta äänestä, jolloin satunnaisuutta ilmaantuu vain hyvin rajallisesti.

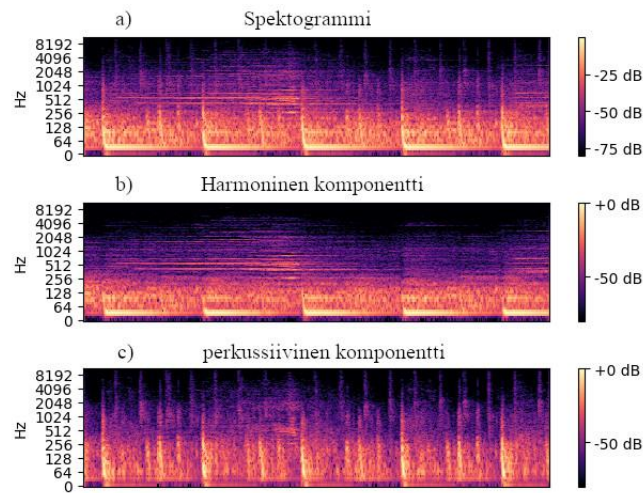
Tutkielman alussa, kolmannessa luvussa käydään läpi yksityiskohtaisesti HPSS-menetelmän periaatteet ja esitetään yksi konkreettinen esimerkki, miten harmonisen ja perkussiivisen komponentin erottelun äänisignaalista voi tehdä. Kolmannessa luvussa perehdytään myös HPSS-menetelmälle oleellisiin työkaluihin: Fourier-muunnokseen ja spektogrammiin. Neljännessä luvussa käydään läpi erilaisia käytännön sovelluksia, joissa on hyödynnetty harmonisen ja perkussiivisen äänen erottelumenetelmiä. Tutkielman lopussa pohditaan HPSS-menetelmän hyödyllisyyttä ja sitä, säilyykö se tärkeänä työkaluna edelleen neuroverkkojen vallatessa alaa musiikin analysoimisessa. Lisäksi mietitään miten HPSS-menetelmä voisi olla hyödyllinen ”sampler-instrumentin” kehittämisessä.

## **2 Tutkimusmenetelmä**

Tutkimusmenetelmänä on käytetty kirjallisuuskatsausta. Katsausta varten on etsitty erityisesti HPSS-menetelmiä käsitteleviä artikkeleita ja kirjallisuutta. Aineisto etsittiin Tampereen Yliopiston Andor-teitokannasta ja IEEEExplore-tietokannasta hakusanoilla: ”Harmonic/percussive sound separation”, ”Harmonic percussive sound separation AND neural network”, ”fourier transform AND audio” ja ”Median filter”. Tietoa HPSS-menetelmästä on etsitty pääosin alan tieteellisistä artikkeleista. Menetelmään liittyvistä oleellisista työkaluista, kuten mediaanisuodattimesta tai Fourier-muunnoksesta tietoa etsittiin alan kirjoista ja luotettavista internetlähteistä. Lähteiden valinnassa on pyritty käyttämään kaikista uusimpia tutkimustuloksia. Lähteitä etsiessä on kuitenkin huomattu, että neuroverkkojen ja BSS-menetelmien (Blind Source Separation) käyttö musikaalisten signaalien analysoimisessa on tuoreemmassa tutkimuksessa yhä yleisempää ja monet HPSS-menetelmää käsittelevät artikkelit ovat jopa 10 vuotta vanhoja. HPSS-menetelmästä löytyy kuitenkin tuoretta tutkimustietoa, joten tutkiminen on edelleen ajankohtaista.

### 3 HPSS-menetelmästä

HPSS- menetelmän tavoitteena on erottaa alkuperäisestä signaalista  $w(t)$  sen harmoninen  $h(t)$  ja perkussiivinen komponentti  $p(t)$ , jossa  $t$  kuvaa aikaa. Ennen kuin varsinaista erottelua voidaan alkaa tehdä, pitää signaali siirtää niin sanotulta aikatasolta spektritasonlle käyttäen Fourier-muunnosta. Erona aikatason ja spektritason välillä on se, että kun aikataso kuvaa signaalin muutosta ajan suhteen, niin spektritaso kuvaa sitä, kuinka paljon signaalista sijaitsee kullakin taajuusalueella milläkin ajan hetkellä. (Broughton, 2008) Fourier-muunnos taas on matemaattinen työkalu, jolla voidaan erottaa signaali sen eri taajuuksiin, joista se pohjimmiltaan koostuu. Tätä työkalua käyttäen voimme koostaa spektrogrammin (Kuva 1, a), joka on välttämätön HPSS-menetelmässä. Olennainen ero perkussiivisella ja harmonisella komponentilla on se, että harmoninen komponentti näyttyytyy spektrogrammissa vaakasuuntaisina pitkinä viivoina (Kuva 1, b) ja perkussiivinen komponentti taas alareunasta yläreunaan menevinä pystysuorina viivoina (Kuva 1, c). (Tachibana ym., 2012.)



Kuva 1 HPSS-menetelmän spektrogrammit. (a) on lähdesignaali  $w(t)$ . (b) on sen arvioitu harmoninen  $h(t)$  komponentti ja (c) arvioitu perkussiivinen komponentti  $p(t)$ .

Onnistuneelle erotukselle on seuraavat ehdot:

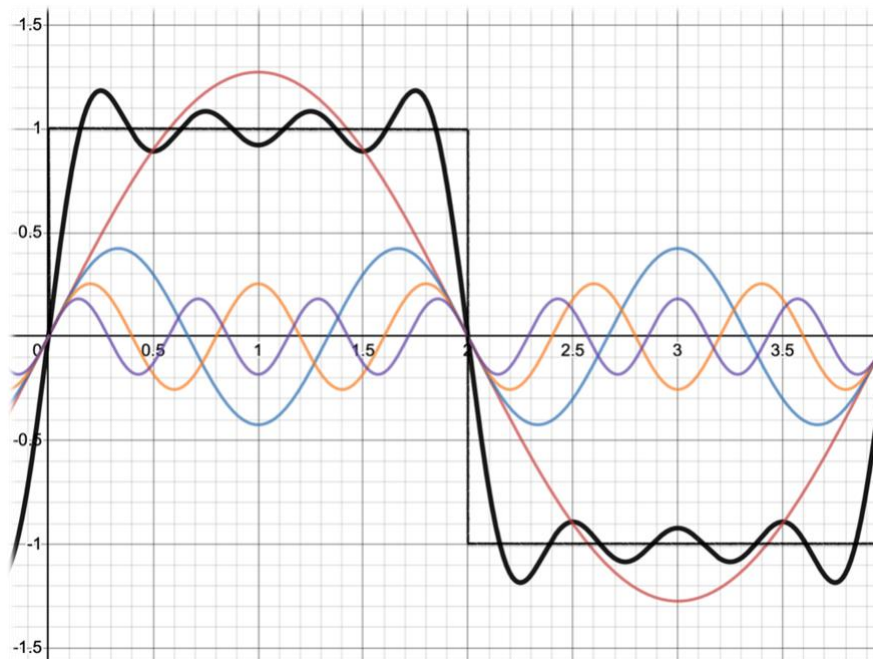
1.  $h(t)$  pitää olla jatkuva ajassa,  $p(t)$  taajuudessa
2.  $h(t) + p(t) \approx w(t)$
3. Jokaisen spektrogrammin komponentin on oltava epänegatiivinen.

### 3.1 Fourier-muunnos

Harmonisen äänisignaalin voidaan ajatella koostuvan yhdestä tai useammasta taajuudesta. Äänellä on usein perusaalto, joka määrittää signaalin kuultavan sävelkorkeuden. Tämän lisäksi se sisältää usein myös osääniä, jotka määrittävät äänenväriä ja ovat perusaallon kerrannaisia. Yksinkertaisin äänen ilmentymä sisältää vain perusaallon, eikä lainkaan kerrannaisia. Tällainen ääni näkyy aikatasolla siniaaltona. Puhtaasta siniaallosta ei voida Fourier-muunnoksella löytää yhtään komponenttia, koska siniaalto esittää jo itsessään vain yhtä taajuutta, jolla ei ole kerrannaisia.

Kuvasta 3 nähdään, miten kanttiaalto rakentuu usean eri siniaallon summasta. Samaan analyysiin pohjautuen myös monimutkaisempi lähdesignaali voidaan erotella sen osääniin ja löytää signaalissa esiintyvät eri taajuudet ja niiden amplitudit. Mikäli signaalista ei ole havaittavissa selkeää perusaaltoa eikä sen osääniä, luokitellaan signaali transientiksi, jonka voi ajatella ilmentävän perkussiivisempaa signaalia. (Watkinson, 2002)

Signaali voidaan esittää aikatasolla, jolla kuvataan signaalin voimakkuuden muutosta ajan funktiona tai spektritasolla, jolla voidaan nähdä signaalin taajuusspektri ajan funktiona. Fourier-muunnoksen avulla signaali voidaan matemaattisesti siirtää aikatasolta spektritasolle tai päinvastoin. (Watkinson, 2002)



Kuva 2 Kanttiaallon Fourier-sarjan neljä ensimmäistä termiä. Musta käyrä esittää neljän ensimmäisen termin summaa. Perusaäni punaisella ja seuraavat kolme osääntä eri väreillä.



Fourier-muunnoksen havainnollistamiseksi seuraavaksi tekstissä käydään läpi Fourier-muunnoksen matemaattinen perusta. Fourier-muunnoksen perimmäinen idea on siinä, että voidaanko jokin funktio  $S(x)$  esittää funktiolla  $f(x)$ , joka on erilaisten sinien ja kosinien summa.

Olkoon  $S(x)$  kanttiaalto ja  $f(x)$  sen fourier-muunnos. Koska kanttiaalto  $S(x)$  voidaan ajatella jaksollisena signaalina, voidaan siihen soveltaa diskreettiä Fourier-muunnosta (DFT) ja esittää se äärellisenä Fourier-sarjana. Alla kanttiaallon  $S(x)$  yhtälö, jossa  $A$  on puolikas amplitudi,  $T$  on jakso:

$$S(x) = A \frac{2i}{\pi} \left[ \tanh^{-1} \left( e^{-\frac{i\pi x}{T}} \right) - \tanh^{-1} \left( e^{\frac{i\pi x}{T}} \right) \right] \quad (1)$$

Fourier-sarja  $f(x)$  yleisesti on:

$$f(x) = a_0 + \sum_{n=1}^{\infty} \left( a_n \cos \frac{n\pi x}{L} + b_n \sin \frac{n\pi x}{L} \right) \quad (2)$$

Koska kanttiaalto on pariton yhtälö niin  $a_0 = a_n = 0$ , joten jäljelle jää yleisen esityksen (yhtälö 2) jälkimmäinen termi.

$$f(x) = \sum_{n=1}^{\infty} \left( b_n \sin \frac{n\pi x}{L} \right) \quad (3)$$

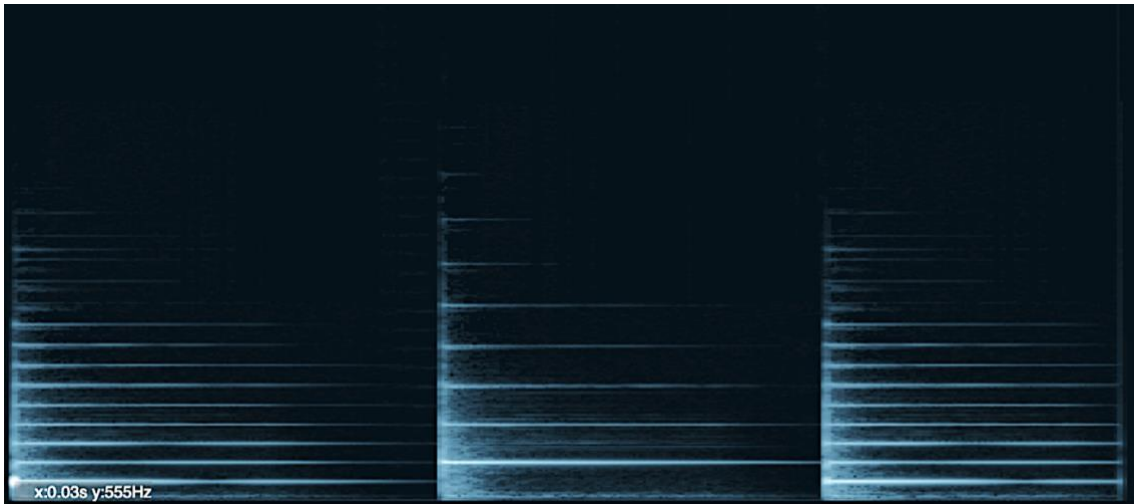
Olkoon kanttiaallon jakso  $2L$  ja puolikas amplitudi  $A = 1$ , tällöin kanttiaallon Fourier-sarjaesitys (yhtälö 3) on:

$$f(x) = \frac{4}{\pi} \sum_{n=1,3,5,\dots}^{\infty} \frac{1}{n} \sin \left( \frac{n\pi x}{L} \right) \quad (4)$$

Tämän esimerkin tarkoituksena oli havainnollistaa matemaattisella tasolla sitä, mitä Fourier-muunnoksessa tapahtuu. Yhtälössä 4 ensimmäinen siniaalto, kun  $n = 1$ , on perusaalto ja loput sinikomponentit osaaäni. Kuvasta 2 voi nähdä miten summa-aalto lähenee kanttiaallon muotoa. (Wolfram) Diskreetti Fourier-muunnos (DFT, "Discrete Fourier transform) ei kuitenkaan ole tarpeeksi tehokas reaaliaikaisen datan käsittelyyn, joten tähän tarpeeseen keksittiin "Fast Fourier transformation" eli FFT. FFT:n avulla laskennan asympotoottiseksi tehokkuudeksi saadaan hieman algoritmista riippuen noin  $O(N \log_2 N)$ . FFT on yksi maailman merkittävimmistä algoritmeista, sillä sen on

mahdollistanut muun muassa nykyäänkin käytettävät videon ja äänen pakkausalgoritmit ja hyvälaatuiset TV-lähetykset. Tämän lisäksi FFT:tä käytetään usein, kun digitaalisesti halutaan suorittaa Fourier-muunnos. (Rao, 2010)

### 3.2 Spektogrammi



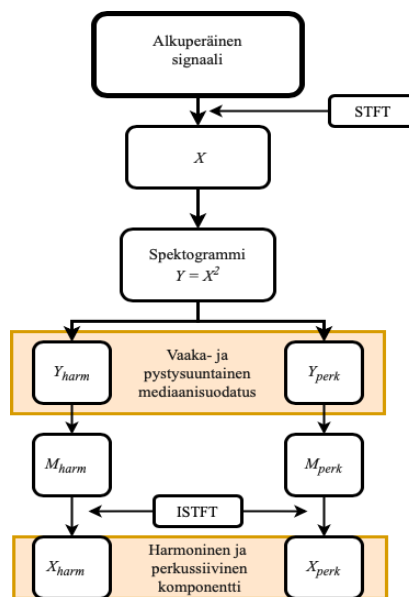
Kuva 3 Spektogrammi. 3 pianon c-nuottia, joista keskimmäinen oktaavia ylempää soitettuna. Y-akselilla on logaritminen esitys taajuudesta ja x-akselilla lineaarinen esitys ajasta. Väriin kirkkaus esittää kyseisen taajuuden voimakkuutta.

Spektogrammi käytetään usein äänen visuaalisena representaationa, jossa yleensä esitetään signaalin taajuus ajan funktiona. Yleensä taajuus esitetään y-akselilla ja aika x-akselilla. Tärkein ominaisuus spektogrammissa on niiden kyky esittää eri taajuusalueiden voimakkuus signaalissa. (Hugill, 2012) Taajuusalueen voimakkuus esitetään kuvassa 3 kirkkaammalla värillä. Mustilla alueilla ei esiinny taajuuksia. Kuvasta 3 katsottaessa yksittäisen pianosävelen alin vaakasuuntainen viiva esittää perusaaltoa ja sen yläpuolella olevat osääniä. Pianosävelien perkussiiviset ominaisuudet ovat myös nähtävissä spektogrammissa, nuotin alussa olevina pystysuorina viivoina, jotka kattavat lähes koko taajuusalueen.

Spektogrammin avulla yleensä esitetään äänidataa, joka on muutettu STFT:llä (Short-Time-Fourier-Transfer) matriisimuotoon. Tämä mahdollistaa monien algoritmien käytön, joita usein käytetään kuvien käsittelyyn. (AaltoWiki) Aikaisemmassa tutkimuksessa on kokeiltu ”tyylinsiirtoneuroverkkoja” (Image neural style transfer networks) spektogrammiin tavoitteena luoda erilaisten äänien yhdistelmiä. (Liu, 2019)

### 3.3 Harmonisen ja perkussiivisen signaalin erottelu

HPSS- menetelmän tavoitteena on erotella äänisignaalin harmoninen ja perkussiivinen komponentti toisistaan. Fourier-muunnos on oleellinen osa menetelmää, koska äänidata on siirrettävä aikatasolta spektritasolle. Nopeamman datankäsittelyn saavuttamiseksi Fourier-muunnokseen on kehitetty erilaisia menetelmiä. Yksi näistä menetelmistä on STFT. STFT:ssä käytetään diskreettiä Fourier-muunnosta ja ikkunafunktiota, joka rajaa äänidatasta vain pienen osan käsiteltäväksi kerralla. (AaltoWiki)



Kuva 4 Perkusiivisen ja harmonisen signaalin erottelumenetelmä mediaanifilttereitä käyttäen.

Aikaisemmassa tutkimuksessa on esitetty miten STFT:n jälkeen, harmonisen ja perkussiivisen komponenttien erottelu voidaan tehdä käyttäen mediaanisuo-dattimia. Kuvasta 4 nähdään, millaisella menetelmällä Solekhanin tutkimusryhmä muokkasi gamelan soittimilla soitettun äänitteen perkussiivisen osan impulsseja erottelemalla signaalin harmonisen ja perkussiivisen komponentin.

Kuvaa 4 seuraten ensin lasketaan alkuperäisen signaalin STFT  $X(m, k)$ . Signaali on nyt siirretty spektritasolle ja sen data on matriisimuotoisena. Tämä matriisi korotetaan toiseen potenssiin ja saadaan signaalin spektogrammi  $Y(m, k) = (X(m, k))^2$ . Tämän jälkeen käytetään spektogrammiin sekä vaakasuuntainen että pystysuuntainen mediaanisuo-datin, joiden avulla saadaan eroteltua harmoninen (vaakasuuntainen) ja perkussiivinen (pystysuuntainen) komponentti. (Solekhan ym., 2019)

Mediaanisuo-dattimen toiminta perustuu siihen, että se laskee jokaiselle signaalin näytteelle tai kuvan pikselille mediaanin kyseisen näytteen tai pikselin naapurinäytteen

tai -pikselien avulla. Laskemalla arvoksi mediaani keskiarvon sijaan signaali sumentuu vähemmän ja lopputulos on tarkempi. (OxFoRef) Mediaanisuolettimia käytetään myös usein poistamaan melua valokuvista. Mediaanisuolettimen muita etuja on sen kyky vähentää tehokkaasti iskumelua ja säilyttää hyvin alkuperäisen kuvan muotoviivan eli ”contour”. (Fernandez-Maloigne ym., 2021)

Mediaanisuolettamisen jälkeen luotujen uusien matriisien, vaakasuuntaisen suolettamisen  $Y_{harm}(m, k)$  ja pystysuuntaisen  $Y_{perk}(m, k)$ , avulla tehdään binäärimaskit  $M_{perk}(m, k)$  ja  $M_{harm}(m, k)$  (Yhtälöt 5 ja 6).

$$M_{perk}(m, k) = \begin{cases} 1, & Y_{harm}(m, k) \geq Y_{perk}(m, k) \\ 0, & \text{muu} \end{cases} \quad (5)$$

$$M_{harm}(m, k) = \begin{cases} 1, & Y_{perk}(m, k) \geq Y_{harm}(m, k) \\ 0, & \text{muu} \end{cases} \quad (6)$$

Nyt binäärimaskien avulla voidaan eristää alkuperäisestä STFT-muunnoksesta  $X(m, k)$  sen harmoninen ja perkussiivinen komponentti.

$$X_{harm} = X(m, k)M_{harm}(m, k) \quad (7)$$

$$X_{perk} = X(m, k)M_{perk}(m, k) \quad (8)$$

Viimeiseksi harmoniselle ja perkussiiviselle komponentille suoritetaan käänteinen Fourier-muunnos eli tässä tapauksessa ISTFT (Inverse-Short-Time-Fourier-Transform). Nyt komponentit ovat siirretty spektritasolta takaisin aikatasolle. Tässä vaiheessa komponentteja voidaan muokata ennen kuin ne yhdistetään takaisin yhteen. (Solekhan ym., 2019)

Solekhanin ryhmän tutkimuksesta ei selvinnyt, kuinka pitkä laskenta-aika menetelmällä oli, mutta STFT:n käyttö viittaisi melko tehokkaaseen menetelmään. Toisessa tutkimuksessa oleva reaaliaikainen HPSS-menetelmä perustui myös STFT:n käyttöön. (Ras ym., 2010) Vaikka mediaanisuolettamia käyttäen ei reaaliaikainen äänidatan käsittely olisikaan mahdollista, menetelmä voisi siltikin olla hyödyllinen sampler-instrumentin toiminnallisuudessa. Mahdollisuus muokata instrumenttiin syötetyn äänitiedoston perkussiivisiä ja harmonisia ominaisuuksia suoraan liitännäisestä olisi vähintäänkin hyödyllistä. On kuitenkin hyvä tiedostaa, ettei mediaanisuolettamien

käyttö ole ainoa keino erotella harmonisia ja perkussiivisia komponentteja toisistaan. Aikaisemmassa tutkimuksessa on kuitenkin todettu mediaanisuo datin pohjaisen menetelmän antavan lähes yhtä hyviä tuloksia, kuin monet laskennallisesti raskaammat ja monimutkaisemmat menetelmät. (Park ym. 2017)

## **4 Sovellukset**

HPSS-menetelmä ei itsessään tarjoa välttämättä mitään selkeää sovelluskohdetta vaan se on enemmänkin työkalu, joka mahdollistaa varsinkin erilaisten MIR-sovelluksien (Musical information retrieval) toteuttamisen. (Ras ym., 2010) HPSS-menetelmää on kuitenkin käytetty muuhunkin tarkoitukseen kuin musiikin analysoimiseen, kuten sydän- ja hengitysäänen mittaamiseen. (Torii ym., 2019) HPSS-menetelmää on käytetty myös sisäkorvaistutteen käyttäjien musiikinkuuntelukokemuksen parantamiseen (Lentz ym., 2020) ja puheen sävyjen analysoimiseen. (Espí ym., 2010)

### **4.1 MIR-sovellukset**

#### *4.1.1 Sointutunnistus*

Suurin osa sovelluksista liittyvät tietysti musiikkiin tavalla tai toisella. Ras ja Wierzchowska kirjassaan ”Advances in Music Information Retrieval” esittävät reaaliaikaisen HPSS- menetelmän ja esittävät ohjelman, jonka he ovat luoneet menetelmänsä pohjalta.

Merkittävin askel sointutunnistuksessa otettiin jo vuonna 1999, kun japanilainen tutkimusryhmä (Kawakami ym., 1999) käytti Markovin piilomallia HMM (Hidden Markov Model) sointukiertojen mallintamiseen niin, että sointukierron ajateltiin edustavan piilossa olevia siirtymätodennäköisyyksiä. HMM-mallin käyttäminen on edelleenkin alalla hyvin yleisesti hyödynnetty sointujen tunnistuksessa, mutta sen tarkkuutta on kyetty huomattavasti parantamaan hyödyntämällä HPSS-menetelmää sen rinnalla. Aikaisemmassa tutkimuksessa saatujen tulosten perusteella ilman HPSS-menetelmän käyttöä, vain HMM-pohjaista tunnistusta käyttäen, sointujen tunnistustarkkuus 180 kappaleen otoksesta oli 45,73 %. Mikäli HPSS-menetelmää hyödyntäen kappaleiden perkussiivista komponenttia vaimennettiin ennen HMM-pohjaista tunnistusta, saatiin tarkkuudeksi 72,48 %. (Ras ym., 2010) Valitettavasti Kawakamin alkuperäistä tutkimusta ei löytynyt tietokannasta, joten sointutunnistuksen historian osalta on luotettava kirjan tekstiin.

#### 4.1.2 Melodian eristys

HPSS-menetelmän hyödyllisyyttä on myös tutkittu melodian eristyksessä. Aikaisemmassa tutkimuksessa pyrittiin eristämään melodia Filippiiniläisestä kansanmusiikista, joka sisälsi erilaisia perkussioita ja melodialinjan joko ihmisen laulamana tai soittimen soittamana. Ensin kappaletta suodatettiin niin, että esimerkiksi laulun tapauksessa, sellaiset taajuudet, joilla ihmisääni todennäköisimmin esiintyy, jätettiin äänisignaaliin. Tämän jälkeen signaalista erotettiin sen harmoninen ja perkussiivinen elementti. Harmonisesta komponentista arvioitiin ”Salience” -funktioita käyttäen voimakkaimmin soivat taajuusalueet ja niiden perusteella luotiin arvio melodiasta.

Tutkijat kokeilivat myös toista tapaa, joka HPSS-menetelmän sijaan käytti SNMF-menetelmää (Shifted non-negative matrix factorization). SNMF-menetelmä erottelee ”sokeasti” signaalin eri taajuusalueilla esiintyvät komponentit sen enempiä erottelematta niitä perkussiivisiin tai harmonisiin komponentteihin. Tutkimuksen lopputuloksena oli, että HPSS-menetelmän käyttö tuotti parempia ennusteita melodiasta. (Disuanco ym., 2015)

#### 4.1.3 Äänenväriin muokkaus ja korvaaminen

Äänenväriillä tarkoitetaan äänilähteen taajuusominaisuuksia, jotka antavat äänelle sen ominaisen soinnin. Äänenväriillä kuvataan usein sitä, miltä jokin kuulostaa. Aikaisemmassa tutkimuksessa esitettiin menetelmä, jolla musiikkikappaleen perkussiivisten ja harmonisten soittimien tai elementtien äänenväri voidaan korvata jonkun toisen soittimen äänenväriä. Käytännössä tämä tarkoittaisi vaikka lauluäänen muuttamista joksikin soittimeksi tai toiseksi lauluääneksi tai bassorummun muuttamista bongorummuksi. Tutkimuksessa käytettiin HPSS-menetelmää harmonisen ja perkussiivisen komponentin eristämiseen, STFT:tä, NMF:ää (Non-negative matrix factorization) ja eri spektrikomponenttien synteisiä. (Nakamura ym., 2014)

Harmonisten äänenvärien muokkauksessa tutkijat jäljittivät alkuperäisen signaalin ja korvaavan signaalin taajuus alueen piikit ja kuopat, ja loivat niistä ”envelopet” eli saatetiedostot. Nämä envelopet antavat suurpiirteisen kuvan kyseisten signaalien sisältämistä taajuus alueista. Tämän jälkeen muodostettiin synteesisignaali, jossa alkuperäisen signaalin envelopea muokattiin muistuttamaan korvaavan signaalin envelopea. Tällä tavoin saatiin signaali, jossa alkuperäisen signaalin spektri on muokattu noudattamaan korvaavaan signaalin spektriä.

Perkussiivisen signaalin muokkauksessa, samoin kuin harmonisen, ensin signaalit eroteltiin HPSS-menetelmällä, jonka jälkeen eri rumpuäänet eroteltiin ”sokealla” erottelulla käyttäen NMF-menetelmää. HPSS-menetelmällä ei voi erotella kuin

harmonisia ja perkussiivisia komponentteja, joten erilaisten perkussiivisten äänien erittelyssä on käytettävä muita menetelmiä, kuten NMF-menetelmää.

Kun perkussiiviset äänet oli eroteltu, käytettiin kahta eri tapaa, muokata alkuperäinen rumpuääni joksikin toiseksi ääneksi. Ensimmäisessä tavassa alkuperäisen signaalin taajuudet korjattiin suoraan muistuttamaan korvaavan signaalin taajuuksia. Tässä menetelmässä kuitenkin haittavaikutuksena syntyi paljon epätoivottua melua, mikäli alkuperäisessä signaalissa jotain heikosti esiintyvää taajuutta, jouduttiin korostamaan liikaa, jotta se saataisiin samalle tasolle kuin korvaavassa signaalissa. Toista tapaa tutkijat kutsuivat ”copy and paste” -menetelmäksi, jossa alkuperäinen signaali pyrittiin korvaamaan suoraan korvaus signaalilla kuitenkin niin, että tämä uusi signaali noudattaisi alkuperäisen signaalin aikamuutoksia.

Yhdeksän koehenkilöä sai arvioida tuloksia asteikolla 1-5 niin, että 1 = äänenvärin muutos ei onnistunut ja 5 = äänenvärin muutos onnistui täydellisesti. Perkussiivisten signaalien korvauksessa keskiarvoksi saatiin n. 2,8 ja harmonisessa korvauksessa n. 2,5. (Nakamura ym., 2014)

## 4.2 Muita sovelluksia

### 4.2.1 Sydän- ja hengitysäänen mittaaminen

Aikaisemmassa tutkimuksessa on myös käytetty harmonisen ja perkussiivisen signaalin määritelmää luovasti. Tutkimuksessa etsittiin kuluttajatason sovellusta, jolla voitaisiin mitata sydän- ja hengitysäntä. Tämän tärkeyttä perusteltiin sillä, että samanaikaisesti mitatuista äänistä voidaan päätellä, mikäli ihminen kärsii sydän- ja verisuonitaukeista tai keuhkosairauksista. Tutkijat mittasivat akustisella sensorilla henkilön kaulasta, korvan alapuolelta, äänisignaalin. Sydämen ajateltiin edustavan tämän signaalin perkussiivista komponenttia ja hengitysäänen sen harmonista komponenttia. Vaikka hengitysäni ei sinänsä esiinny millään tietyllä taajuusalueella, se kuitenkin edustaa pidempää ajassa tapahtuvaa vaakasuuntaista signaalia, kun taas sydämensyke sopii paremmin perkussiivisen signaalin määritelmään.

Tutkijat käyttivät HPSS-menetelmää mediaanisuodattimia hyödyntäen. Binäärimaskien sijaan tutkijat käyttivät Wiener-suodatinta, jolla he loivat pehmeän maskin harmonisen ja perkussiivisen komponentin erotteluun. Tutkijat onnistuivat mittaamaan tällä tavalla sydämen sykkeen riittävän tarkasti ja lisäksi vähentämään sydänäänien vaikutusta hengitysänessä. Sydänäänien vaikutusta ei pystytty kuitenkaan poistamaan kokonaan hengitysäneestä. Tutkimuksessa käy ilmi, että menetelmällä hengitysänen mittaustarkkuutta saatiin parannettua verrattuna siihen, että HPSS-menetelmää ei olisi käytetty, mutta ei oteta kantaa, että onko mittaustarkkuus riittävän tarkka käytännön sovellukseen. (Torii ym., 2019)

#### *4.2.2 Sisäkorvaistutteen käyttäjien musiikinkuuntelukokemuksen parantaminen*

Aikaisemmassa tutkimuksessa on etsitty ratkaisua sisäkorvaistutteen käyttäjien musiikinkuuntelukokemukseen. Sisäkorvaistutteen on todettu parantavan huomattavasti käyttäjien kykyä saada selvää puheesta, mutta musiikin kuuntelu on ilmaistu olevan jopa epämiellyttävää. Tämä johtuu siitä, että istute ei kykene tuottamaan tarpeeksi tarkkaa spektriä, jolla musiikin eri taajuusalueista saisi selvää. Tutkimuksessa käytettiin HPSS-menetelmää mediaanifilttereitä hyödyntäen erottelemaan musiikkikappaleen harmoniset ja perkussiiviset elementit toisistaan. Tämän jälkeen harmonisen komponentin voimakkuutta säädettiin. Tärkeimmän melodisen elementin ajateltiin näyttäytyvän harmonisessa komponentissa kaikkein voimakkaimpana, joten se säilytettiin ennallaan ja muiden taajuusalueiden voimakkuutta vähennettiin. Tämän lisäksi perkussiivisen komponentin voimakkuutta saatettiin joissain tapauksissa korostaa entisestään.

Sisäkorvaistutteen käyttäjät raportoivat tämän menetelmän parantavan kokonaisuuntelukokemusta. Tutkijat huomasivat kuitenkin, että menetelmää ei voi soveltaa kaikkiin musiikkigenreihin. Esimerkiksi klassisen musiikin kuuntelukokemusta ei pystytty menetelmällä parantamaan. Tämän johtui siitä, että klassinen musiikki ei usein sisällä selkeää rymistä pulssia, jota korostamalla kappaleesta saisi paremmin otteen. Myös harmonisesti klassinen musiikki on huomattavasti monimutkaisempaa, eikä aina sisällä selkeää päämelodiaa. (Lentz ym., 2020)

#### *4.2.3 Puheensävyyden tunnistaminen*

Puheentunnistuksessa ja sen analysoimisessa on tärkeää, että signaalista, jota käsitellään saadaan mahdollisimman tarkasti eroteltua se osa, joka sisältää puhetta. Tätä aluetta käsitellään puheaktiivisuuden ilmaisun tutkimuksessa (VAD, Voice activity detection). Eräs suuri ongelma puheaktiivisuuden ilmaisun tutkimuksessa on puhetta sisältävien signaalien erottelu äänitteestä, joka sisältää paljon melua. Melua äänitteeseen syntyy usein, mikäli puhe on äänitetty liian hiljaa ja mahdollisesti huonolla mikrofonilla. VAD-menetelmien tavoitteena on eristää puhetta sisältävä osa signaalista, jotta sitä voitaisiin analysoida.

Aikaisemmassa tutkimuksessa ehdotettiin, että HPSS-menetelmällä voitaisiin yrittää erottaa puhetta sisältävästä signaalista sen harmoninen komponentti ja tällä tavalla saada signaalista tunnistettua sen puhetta sisältävät osat. Tutkijat vertailivat 5 eri menetelmää puheaktiivisuuden ilmaisuun. HPSS-menetelmä osoittautui tehokkaaksi ja toimivaksi työkaluksi. HPSS-menetelmä toimi myös hyvin olosuhteissa, joissa alkuperäinen signaali sisälsi vaihtelevaa melua. (Espil ym., 2010)



## 5 Pohdinta

HPSS-menetelmä itsessään ei tarjoa ratkaisua mihinkään tiettyyn ongelmaan, vaan se on pikemminkin nopea ja hyödyllinen työkalu musiikin analysoimiseen: Usein HPSS-menetelmää käytetäänkin signaalin esivalmistelussa yhdistelmänä jonkin toisen menetelmän kanssa. Sointujentunnistuksessa menetelmällä vaimennettiin perkussiivista signaalia, jotta HMM-mallilla kyettiin tuottamaan parempia ennusteita soinnuista (Ras ym., 2010) ja ennen melodian eristystä signaalista eroteltiin sen harmoninen komponentti. (Disuanco ym., 2015) HPSS-menetelmää käytettiin myös yhdistelmänä ”sokean” erottelun NMF-menetelmän kanssa, jolloin erillisten rumpuäänien erottelu toisistaan saatiin tehtyä tarkemmin, mitä se olisi onnistunut, mikäli harmoninen komponentti olisi ollut edelleen osa signaalia. (Nakamura ym., 2014)

Vaikka HPSS-menetelmästä on erilaisia toteutuksia, tässä tutkielmassakin esitetty mediaanisuodattimia ja STFT:tä käyttävä malli on laskennallisesti kevyimmästä päästä tuottaen kuitenkin hyviä tuloksia. (Park ym. 2017) Onkin tärkeää, että niinkin monimutkaisen datan kuin musiikki, analysoimiseen on laskennallisesti nopeita ratkaisuja, joilla saada eriytettyä signaalista oleellisia ominaisuuksia analysoitavaksi.

Järjestelemättömän datan määrä maailmassa on kasvanut viimeisten vuosikymmenien aika valtavasti. Musiikki kuuluu tähän kategoriaan. Tämän datan järjestely, luokittelu ja ryhmittely onkin noussut tärkeäksi ongelmaksi. Aikaisemmassa tutkimuksessa neuroverkoilla on pyritty myös analysoimaan musikaalisten signaalien ominaisuuksia. Genre- ja soitintunnistus on yksi tärkeä luokittelu, jota musikaalisille signaaleille pyritään tekemään. HPSS-menetelmälle on löytynyt myös hyötykäyttöä tässä kontekstissa. Tunnistukseen käytetyn neuroverkon tunnistustarkkuutta kyettiin parantamaan 4%, kun signaali eroteltiin ennen genretunnistusyrittystä. (Gwardys ym., 2014). Neuroverkkojen avulla olisi varmasti mahdollista suorittaa vielä tarkempi harmonisen ja perkussiivisen komponentin erottelu kuin ”perinteisillä” HPSS-metodeilla. On kuitenkin mielekästä kysyä, että mikäli on mahdollista saada lähes yhtä hyvä tulos pienemmillä resursseilla ja laskennallisesti yksinkertaisemmin, olisiko tällaisen neuroverkkototeutuksen tekemisestä suurta hyötyä.

”Sampler-instrumenteissa” HPSS-menetelmän hyödyt voisivat olla laajat. Mahdollisuus muokata äänitiedoston spektridataa reaaliaikaisesti tällaisessa instrumentissa, olisi ominaisuus, joka olisi musiikkiteknologian kentällä mullistavaa. Miksi tällaista teknologiaa ei juurikaan ”sampler-instrumenteissa” käytetä, voi tarkoittaa sitä, että erottelun tarkkuus ei ole tarpeeksi hyvä ja tulokset eivät ole miellyttäviä kuunnella. On myös mahdollista, että ala on hidas reagoimaan uusiin menetelmiin tai että kuluttajat eivät vielä ole valmiita näin erikoisille ominaisuuksille. Usein kun ”sampler-instrumentti” on valmistettu, saattaa sitä soittaessa huomata, että äänitteitä pitäisi vielä jotenkin muokata soittoon soveltuvammaksi. Yksi vaihtoehto on tietenkin äänittää kaikki,

mahdollisesti tuhannet, äänet uudestaan ja valmistaa instrumentti uudestaan, mutta tällaisissa tilanteissa jo pelkästään harmonisen ja perkussiivisen komponentin voimakkuuksien säätäminen, olisi erittäin hyödyllinen työkalu. Puhumattakaan siitä, että näiden komponenttien ominaisuuksia itsessään voisi vielä muokata tavalla tai toisella.

HPSS-menetelmästä voisi mahdollisesti olla myös hyötyä, mikäli halutaan omien äänitteiden sijaan, käyttää esimerkiksi jossakin kappaleessa esiintyviä rumpuääniä ”sampler-instrumentin” valmistamiseen. Tällöin HPSS-menetelmällä voitaisiin yrittää eritellä kappaleesta sen perkussiivinen komponentti, josta taas etsiä yksittäiset halutut rumpuäänit. Tällaista ominaisuutta ei markkinoilla olevissa ”sampler-instrumenteissa” ole.

Signaalin siirtäminen aikatasolta spektritasolle on hyvin oleellinen menetelmä musiikkiteknologiassa. On olemassa satoja liitännäisiä, jotka hyödyntävät spektritasoa äänen ominaisuuksien kuvauksessa. Yleisintä on kuitenkin vain visualisoida erilaisia musiikkiin ja ääneen liittyviä ominaisuuksia eikä niinkään muokata näitä ominaisuuksia. On hyvä kuitenkin tiedostaa, että spektritason muokkausta saatetaan monissakin liitännäisissä tehdä, vaikkei sitä käyttäjillä mitenkään erityisesti näytetä.

Tätä tutkielmaa tehdessä on käynyt ilmi, kuinka lähellä äänidata on kuva dataa. Ääni esitetään aikatasossa vektorina, joka sisältää jonkin määrän ääninäytteitä. Fourier-muunnoksen jälkeen, aikatasolta spektritasolle siirryttäessä, äänidata muuttuukin matriisimuotoiseksi, jolloin yksittäisten ääninäytteiden voidaankin ajatella edustavan yhtä pikseliä, jota ne spektrogrammin tapauksessa tekevätkin. Tämä äänidatan ja kuvadatan rajan hämärtyminen onkin ollut tutkimusta tehdessä yksi mielenkiintoisimmista havainnoista.

Tuoreemmassa tutkimuksessa esiintyy myös usein BSS-menetelmiä, joista yhtenä yleisimpänä NMF-menetelmä. Niissä pyritään erottelemaan kaikki erilaiset äänilähteet tarkastelematta erityisesti niiden perkussiivisuutta tai harmonisuutta. Tällaisen menetelmän käyttö musiikkiteknologiassa ja erityisesti ”sampler-instrumenteissa” voisi myös olla jatkotutkimuksen arvoista. Voisiko BSS-menetelmällä olla mahdollista erotella äänitteestä sen kaikki yksittäiset äänet ja käyttää niitä ”sampler-instrumentin” luomisessa. Näin ei tarvitsisi erikseen äänittää soittimen jokaista ääntä, vaan ne voisi erotella yhdestä pidemmästä äänitteestä.

## 6 Yhteenveto

Tutkielman tavoitteena oli selvittää, mihin harmonisen ja perkussiivisen äänen erottelumenetelmä perustuu ja onko sitä mahdollista käyttää reaaliaikaisesti äänen käsittelyyn. Aikaisemman tutkimuksen perusteella luvussa 3 esitetty erottelumenetelmä,

jossa hyödynnettiin mediaanisuodattamia, on yksi laskennallisesti tehokkaimmista tavoista suorittaa erottelu. Mediaanisuodattamien avulla tehty erottelu myös vastasi tuloksen laadultaan monia monimutkaisempia ja raskaampia tapoja. (Park ym. 2017) Reaaliaikaisen erottelumenetelmän vaatimuksina vaikuttaa olevan myös STFT:n käyttö. Alan kirjallisuudessa esitettiin myös yksi mahdollinen sovellus, jossa HPSS-menetelmää käytettiin reaaliaikaisesti äänidatan käsittelyyn, mutta toteutuksesta jäi epäselväksi, käyttikö sovellus jo valmiiksi tiedossa olevaa äänitiedostoa vai kykenikö se myös prosessoimaan reaaliaikaisesti puskureina saapuvaa äänidatavirtaa. (Ras ym., 2010)

Aikaisempaan tutkimukseen vedoten voidaan todeta, että HPSS-menetelmän käyttö ”Sampler-instrumentissa”, olemassa olevien äänitiedostojen harmonisten ja perkussivisten ominaisuuksien muokkaamiseen, on täysin toteutettavissa. Menetelmän soveltuvuudesta reaaliaikaisen äänidatavirran käsittelyyn tarvitaan lisätutkimusta.

## Lähdeluettelo

- AaltoWiki. (2022). <https://wiki.aalto.fi/display/ITSP/Spectrogram+and+the+STFT> (haettu: 4.8.2022)
- Bisharad, D., & Laskar, R. H. (2019). Music genre recognition using convolutional recurrent neural network architecture. *Expert Systems*, 36(4). <https://doi.org/10.1111/exsy.12429>
- Broughton, S.A., & Bryan, K. (2008) *Discrete Fourier Analysis and Wavelets: Applications to Signal and Image Processing*. New York: Wiley. p. 72.
- Disuanco, J., Tan, V., & de Leon, F. (2015). Study of automatic melody extraction methods for Philippine indigenous music. 2015 IEEE International Conference on Control System, Computing and Engineering (ICCSCE), 464–469. <https://doi.org/10.1109/ICCSCE.2015.7482230>
- Espi, M., Miyabe, S., Nishimoto, T., Ono, N., & Sagayama, S. (2010). Analysis on speech characteristics for robust voice activity detection. 2010 IEEE Spoken Language Technology Workshop, 151–156. <https://doi.org/10.1109/SLT.2010.5700838>
- Fernandez-Maloigne, C., Robert-Inacio, F., & Macaire, L. (2012). *Digital color imaging*. London: ISTE.
- Gwardys, G., & Grzywczak, D. (2014). Deep Image Features in Music Information Retrieval. *International Journal of Electronics and Telecommunications*, 60(4), 321–326. <https://doi.org/10.2478/eletel-2014-0042>
- Hugill, A. (2012). *The digital musician*. Taylor & Francis Group.
- Kawakami, T., Nakai, M., Shimodaira, H., Sagayama, S. (1999). Harmonization for melody using HMM. In: Proc. JHES, F-61, p. 361 (1999) (japaniksi)
- Lentz, B., Nagathil, A., Gauer, J., & Martin, R. (2020). Harmonic/Percussive Sound Separation and Spectral Complexity Reduction of Music Signals for Cochlear Implant Listeners. ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 8713–8717. <https://doi.org/10.1109/ICASSP40776.2020.9052920>
- Liu, X., Delany, S.J., McKeever, S. (2019). Sound Transformation: Applying Image Neural Style Transfer Networks to Audio Spectrograms. In: Vento, M., Percannella, G. (eds) *Computer Analysis of Images and Patterns. CAIP 2019. Lecture Notes in Computer Science*, vol 11679. Springer, Cham. [https://doi-org.libproxy.tuni.fi/10.1007/978-3-030-29891-3\\_29](https://doi-org.libproxy.tuni.fi/10.1007/978-3-030-29891-3_29)
- Nakamura, T., Kameoka, H., Yoshii, K., & Goto, M. (2014). Timbre replacement of harmonic and drum components for music audio signals. 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 7470–7474. <https://doi.org/10.1109/ICASSP.2014.6855052>

- OxFoRef.(2022). <https://www-oxfordreference-com.libproxy.tuni.fi/view/10.1093/acref/9780198725725.001.0001/acref-9780198725725-e-2911> (haettu: 4.8.2022)
- Park, J., Shin, J., Lee, K., Park, J., Jaeyoung Shin, & Kyogu Lee. (2017). Exploiting Continuity/Discontinuity of Basis Vectors in Spectrogram Decomposition for Harmonic-Percussive Sound Separation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 25(5), 1061–1074. <https://doi.org/10.1109/TASLP.2017.2681742>
- Rao, Kim, D. N., & Hwang, J. J. (2010). *Fast Fourier Transform - Algorithms and Applications* (1st ed. 2010.). Springer Netherlands. <https://doi.org/10.1007/978-1-4020-6629-0>
- Ras, Z.W., & Wieczorkowska, A. (2010). *Advances in Music Information Retrieval*. Ed. Zbigniew W. Ras and Alicja. Wieczorkowska. 1st ed. 2010. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010. <https://doi-org.libproxy.tuni.fi/10.1007/978-3-642-11674-2>
- Solekhan, S., Suprpto, Y. K., & Wirawan, W. (2019). Impulsive spike enhancement on gamelan audio using harmonic percussive separation. *International Journal of Electrical and Computer Engineering (Malacca, Malacca)*, 9(3), 1700–1710. <https://doi.org/10.11591/ijece.v9i3.pp1700-1710>
- Tachibana,H., Kameoka,H., Ono, N., & Sagayama, S. (2012). "Comparative evaluations of various harmonic/percussive sound separation algorithms based on anisotropic continuity of spectrogram," 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2012, pp. 465-468, doi: 10.1109/ICASSP.2012.6287917.
- Torii, K., Nakashima, S., Nakamura, H., Ooyagi, T., Tanaka, K., & Yagami, H. (2019). Distinction of Heart Sound and Respiratory Sound Using Body Conduction Sound Sensor Based on HPSS. *Proceedings of the 7th ACIS International Conference on Applied Computing and Information Technology (ACIT 2019)*. Association for Computing Machinery, New York, NY, USA, Artikkeli 31, 1–6. <https://doi-org.libproxy.tuni.fi/10.1145/3325291.3325386>
- Watkinson. (2002). *An introduction to digital audio* (2nd ed.). Focal Press.
- Wolfram. (2022). <https://mathworld.wolfram.com/FourierSeriesSquareWave.html> Mathworld. (haettu 28.6.2022)

