Rumpuäänien klusterointi k:n keskipisteen menetelmällä ja itseorganisoituvalla kartalla

Tero Tarvainen
Tero.Tarvainen@helsinki.fi

Klusterointimenetelmät-seminaari
Helsingin yliopisto, tietojenkäsittelytieteen laitos
Raportti C–2002–54, s. 12–20, marraskuu 2002

Tiivistelmä

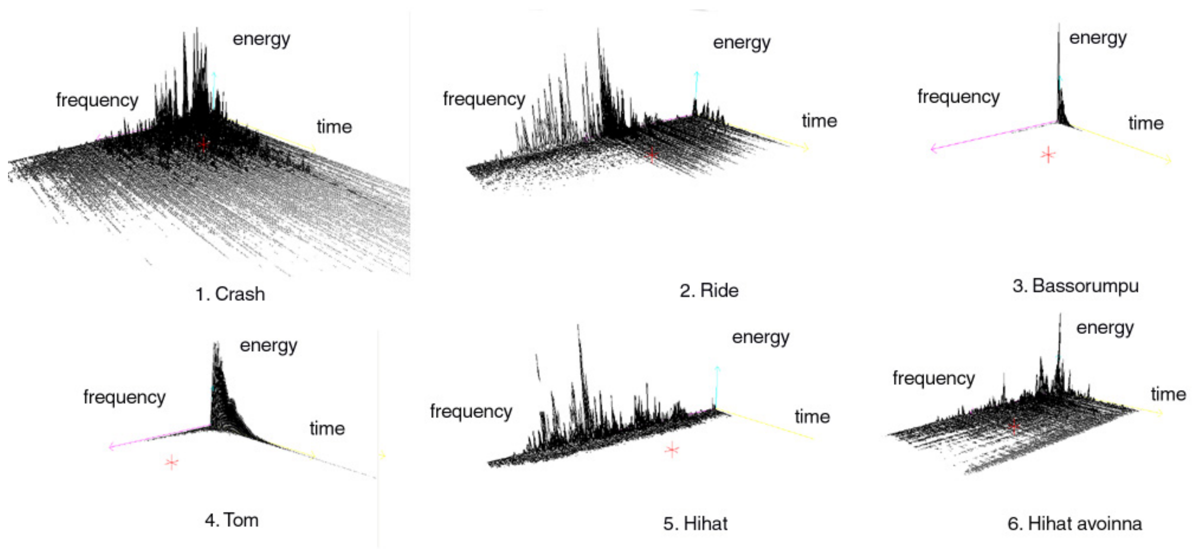
Tämä raportti esittelee toteutetun järjestelmän rumpuäänien luokitukselle. Järjestelmän tarkoituksena on tutkia K:n keskipisteen menetelmän ja itseorganisoituvan kartan sopivuutta rumpuäänien luokitukseen. Sopivien piirteiden etsimistä rumpuäänistä tutkitaan aikaisemmin tehtyjen tutkimusten nojalla. Molempien menetelmän sopivuutta piirrevektorin hyvyyden arvioinnille pohditaan. Ohjelmaa testataan kouluttamalla mallit neljälläsadalla rumpuäänellä, jonka jälkeen koulutetut mallit laitetaan luokittelemaan yli kahtasataa rumpuääntä. Piirrevektorina järjestelmässä käytetään vektoria, joka mallintaa äänen taajuudellista ja ajallista käyttäytymistä.

1 Johdanto

Toteutettu ohjelma luokittelee syötteenä annettuja rumpuääniä. Työn tarkoituksena on vertailla kahden eri klusterointimenetelmän toimivuutta rumpuäänien luokittelussa. Tutkitut menetelmät ovat itseorganisoituva kartta ja k:n keskipisteen klusterointi. Työn toisena motiivina on etsiä joukko piirteitä rumpuäänistä, joilla rumpuäänet voidaan vielä erotella toisistaan, ja jolle joukon koko on mahdollisimman pieni.

Motivaatio, miksi rumpuäänet voidaan luokitella eri luokkiin on se, että niinhän ihminenkin pystyy yleisessä tapauksessa helposti luokittelemaan ne. Toinen vinkki saadaan, kun tarkastellaan rumpuäänien spektrogrammeja. Kuvasta 1 näkyy miten eri rumpuäänet käyttäytyvät eri tavoin. Selvimpiä eroja löytyy äänen dynamiikassa, eli miten ääni syttyy tai sammuu, ja äänen käyttämissä taajuusalueissa. Spektrogrammin tarkastelusta saadaankin vihiä niistä piirteistä, jotka ovat tärkeitä kuvaamaan erilaisia rumpuääniä.

Ohjelman toteutuksessa rajoitutaan tarkastelemaan yleisempien rumpujen muodostamia ääniä. Mahdollisia luokkia toteutetussa järjestelmässä on näin ollen yhteensä kahdeksan: Bassorumpu, virveli, tomrumpu, ride-pelti, hihat-pellit, hihat-pellit avoinna, crash-pelti ja splash-pelti. Klusteroinnissa käytetään studiossa nauhoitettuja puhtaita rumpuääniä. Äänet ovat nauhoitettu 44100 Hz näytteenottotaajuudella ja ne eivät sisällä ylimääräistä melua, kuten kohinaa tai muiden instrumenttien ääntä, vain ainoastaan kyseenomaisen rummun iskun.



Kuva 1: Spectrogrammi crash-pellista, ride-pellistä, bassorummusta, tomrummusta, hihat-pellistä ja hihat-pellistä avoinna.

Myös tavallisesta poikkeavat iskutekniikat, kuten vispilöillä tuotettu ääni, on jätetty pois. Käytetyt äänet on valikoitu siten, että ne kattavat mahdollisimman suuren variaation (sama rumpu eri korkeuksille viritettynä jne.). Äänet on nauhoitettu eri studioissa, eri mikrofoneilla ja erilaisissa tiloissa. Tämä on tärkeää siksi, että se poistaa mahdollisen vaaratilanteen, jossa käytetty malli oppii mikrofonin tai tilan aiheuttamia erityisiä piirteitä.

Opetusdatan hankkiminen oli erittäin vaikeaa. Tästä syystä testidatassa saattaa olla muutama sama rumpuääni kuin opetusdatassakin. Aineisto sisältää myös sellaisia ääniä, joiden tunnistaminen ihmisellekin on miltein mahdotonta. Tälläisiä ääniä ovat esimerkiksi meluisat virvelit ja synteettiset, koneella tehdyt, rumpuäänet.

2 Rumpuäänien vektorisointi

Rumpuäänien vektorisointi on oletettavasti kriittinen osa klusteroinnin onnistumiselle. Halutaan piirrevektori, joka kuvaa ääntä mahdollisimman tarkasti, mutta jonka dimensio on samalla mahdollisimman pieni. Hyvän piirrevektorin löytämiseksi on hyvä tutkia, mitkä ovat ne piirteet audio datassa, jonka mukaan ihminenkin tuntuisi erottelun tekevän. Tässä kappaleessa tarkastellaan niitä piirteitä, jotka ovat tärkeitä rumpuäänien erottelulle.

2.1 Dynamiikka

Ensimmäisenä tärkeänä piirteenä tarkastellaan rumpuäänen dynamiikkaa. Dynamiikalla tarkoitetaan tässä sitä, kuinka nopeasti ääni syttyy ja sammuu (attack, decay). Äänen syttymis- ja sammumiskäyttäytymisellä on todettu olevan suuri merkitys äänien erottelulle:

> "It seems that information about the way an object is hit is conveyed by the attack segment, where as the decay segment conveyes information about the shape and material of the beaten object"[LMe94].

> "Temporal charasteristics of the attack seem to retain most of the information for timbre discrimination"[Gre77].

Yleisesti tutkimuksissa, joissa luokittelaan instrumentteja, tutkittava ääni jaetaan erikseen syttymis- ja sammumisosioihin, joista sitten poimitaan piirteitä [GPD00, HYG02, Ero01]. Syttymis- ja sammumisosion rajan laskemiseksi käytetään erilaisia menetelmiä. Naiivi tapa on etsiä tutkittavasta äänitiedostosta maksimiarvo ja katsoa missä kohdassa maksimiarvo saavutetaan. Tämä menetelmä toimii huonosti mikäli ääni sisältää äkillisiä piikkejä tai jos ääni soi kauan. Hienostuneenpi menetelmä on laskea äänitiedostolle RMS (root means square) - energiataso ja tulkita syttymis/sammumis rajaksi se kohta, jolloin äänen energia saavuttaa ensimmäisen kerran lasketun RMS arvon [Ero01]. Tämä menetelmä sopii hyvin soiville instrumenteille, mutta heikommin lyömäsoittimille.

Dynamiikasta käytettyjä piirteitä ovat yleisesti syttymisaika (aika alusta syttymis-/sammumisrajalle) ja sammumisaika (syttymis-/sammumisrajalta loppuun). Motivaatio tälle rumpuäänien suhteen on se, että yleisesti rummut syttyvät huomattavasti nopeammin kuin pellit ja pellit sammuvat (soivat) kauemmin kuin rummut.

2.2 Taajuuskäyttäytyminen

Toinen tärkeä piirre on äänten taajuuskäyttäytyminen. Taajuuskäyttäytyminen kuvaa äänen täysin ja siten sisältää erittäin paljon informaatiota. Tästä syystä tarkoitus on etsiä pieni määrä piirteitä, jotka erottelevat eri rumpuäänet toisistaan mahdollisimman hyvin. Tutkimalla eri rumpuäänien spektrogrammeja huomataan, että eroja löytyy mm. taajuusalueissa ja siinä miten eri taajuusalueet syttyvät ja sammuvat eri aikoihin.

Yleinen tapa tutkia taajuuskäyttäytymistä on generoida spectrogrammi datasta käyttäen Fourier muunnosta (erityisesti FFT algoritmia). Koska tutkittava data ei ole säännöllistä, tarvitaan ikkunointia Fourier muunnoksen laskemiseksi. Yleisesti ikkunointina käytetään Hanning-ikkunointia (esim. [GoM95]).

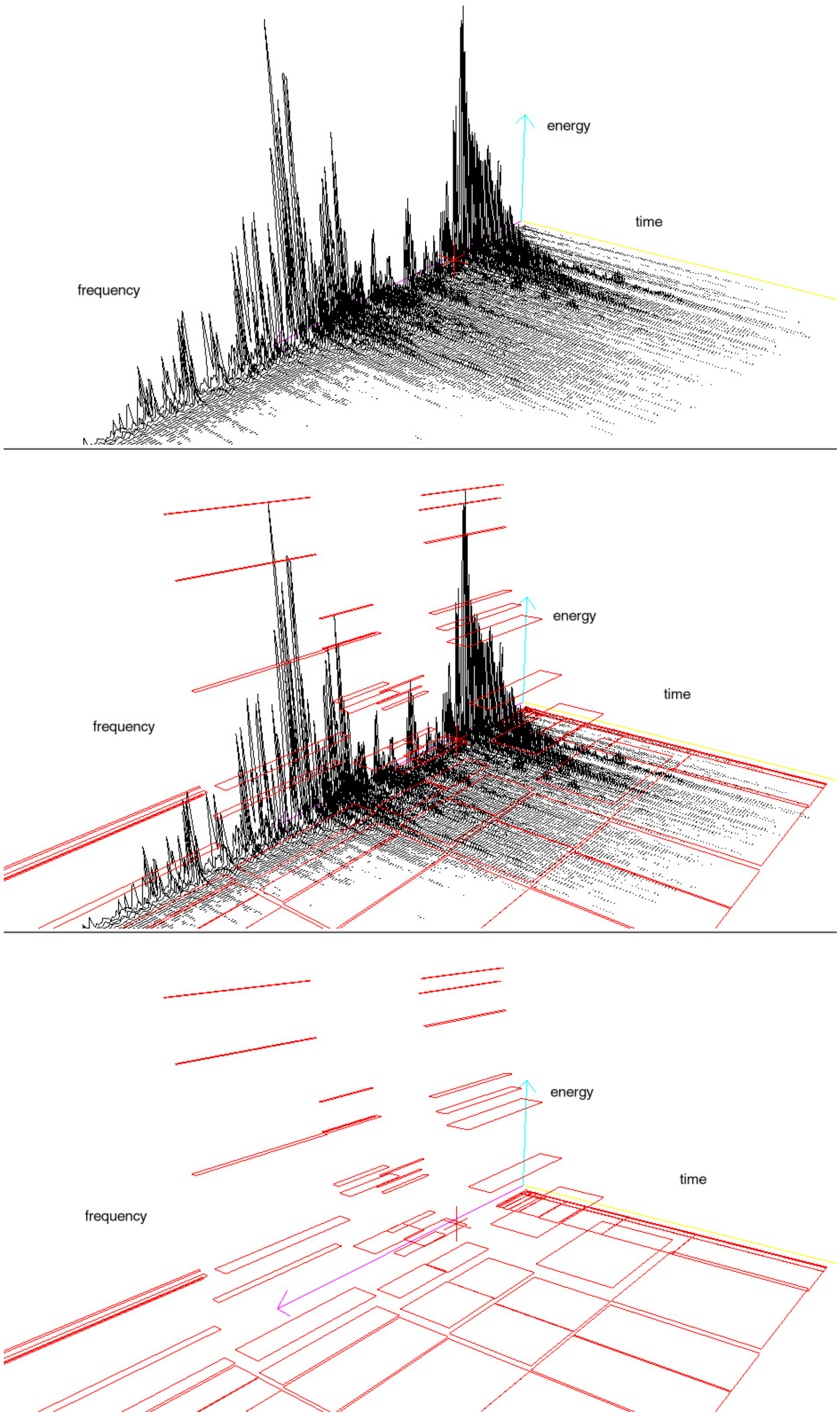
FFT:n lisäksi on kehitetty muita tapoja taajuuskäyttäytymisen tutkimiseen. Erityisesti ihmisen kuuloa simuloivia menetelmiä on kehitelty paljon. Tämä tutkielma ei kuitenkaan syvenny tähän tämän tarkemmin. Hyviä artikkeleita ihmisen kuuloa simuloivista järjestelmistä on esitetty paljon. Näistä esimerkkeinä artikkelit [Lyo82, PRH92, Sla88, Med86].

2.3 Piirrevektori

Toteutetussa ohjelmassa piirrevektoriin lasketaan piirteitä sekä dynaamisesta käyttäytymisestä että taajuuskäyttäytymisestä. Ohjelma tekee Fourier muunnoksen datalle käyttäen 1024 näytteen (23.2 ms) Hamming ikkunaa. Ikkunaa siirretään 256 näytteen verran (5.8 ms) jokaisen Fourier muunnoksen yhteydessä. Käytetyn menetelmän taajuusresoluutio on 43 Hz ja aikaresoluutio 5.8 ms.

Äänen aika- ja taajuuskäyttäytyminen on mallinnettu jakamalla ääni sekä aika- että taajuusalueittain osiin. Kuvassa 2 on esitetty käytetty ositus. Taajuusalue on jaettu yhdeksään ja aika kymmeneen osaan. Yhteensä siis yhdeksänkymmentä taajuus/aika-ikkunaa. Ikkunat ovat vaihtelevan kokoisia. Ikkunan koko on kiinnitetty etukäteen, joten se ei ole riippuvainen tarkasteltavasta äänestä. Taulussa 1 on esitty taajuusalueen jaottelu ja taulussa 2 esitetään ajan jaottelu. Koska äänen syttymisosio sisältää erittäin paljon informaatiota äänestä, ikkunoiden tiheys on suurin äänen alkupuolella.

Tämän hetkisessä toteutuksessa piirrevektori lasketaan seuraavasti. Piirrevektorin ensimmäiseen komponenttiin lasketaan äänen kesto. Toisin sanoen, kuinka kauan ääni kestää. Ää-



Kuva 2: Ääni jaetaan yhdeksäänkymmeneen taajuus/aika-ikkunaan. Ensimmäisessä kuvassa näkyy ainoastaan äänen spektrogrammi. Toiseen kuvaan on lisätty ikkunointi. Kolmanteen kuvaan on jätetty ainoastaan kyseenomaiseen ääneen liittyvät ikkunat. Ikkunoiden korkeus ilmaisee ikkunan arvon.

Taajuusikkuna	*Taajuusalue*
1	0 - 43 Hz
2	44 - 86 Hz
3	87 - 430 Hz
4	431 - 2153 Hz
5	2134 - 6460 Hz
6	6461 - 8613 Hz
7	8614 - 10767 Hz
8	10768 - 15073 Hz
9	15074 - 22050 Hz

Taulukko 1: Ikkunoiden vastaavat taajuusalueet

Aikaikkuna	*Näyteväli*	*Aikaväli*
1	0 - 256	0 - 5.8 ms
2	257 - 512	5.8 - 11.6 ms
3	513 - 1024	11.6 - 23.2 ms
4	1025 - 1792	23.2 - 40.6 ms
5	1793 - 3072	40.6 - 69.7 ms
6	3073 - 5632	69.7 - 127.7 ms
7	5633 - 8192	127.7 - 185.8 ms
8	8193 - 13312	185.8 - 301.9 ms
9	13313 - 20992	301.9 - 476.0 ms
10	20993 - 33792	476.0 - 766.3 ms

Taulukko 2: Ikkunoiden aikavälit

nen katsotaan alkaneen/loppuneen silloin, kun sen energiataso nousee/laskee annetun raja-arvon yli/ali. Raja-arvona käytetään äänen maksimiarvoa jaettuna tuhannella. Saatu kesto äänelle normalisoidaan 0.5 sekunnilla ennen kuin se asetetaan piirrevektorin arvoksi. Saadut arvot ovat yleisesti välillä $[0, 2]$. Äänet, joiden pituus ylittää sekunnin, ovat harvinaisia.

Piirrevektorin toisen komponentin arvoksi asetetaan äänen vahvimman taajuuden indeksi (FFT taulussa). Indeksi on normalisoitu taulun koolla, jolloin indeksin mahdolliset arvot ovat välillä $[0, 1]$.

Seuraavat yhdeksänkymmentä komponenttia piirrevektoriin saadaan taajuus/aika-ikkunoista. Jokaisen ikkunan sisältä etsitään se taajuus, jolla on suurin energia alueen sisällä. Tämä arvo normalisoidaan äänen vahvimman taajuuden maksimiarvolla. Saatu normalisoitu arvo asetetaan ikkunan arvoksi. Näin ikkunoiden arvot ovat välillä $[0, 1]$.

3 Klusterointimenetelmät

Työssä keskitytään kahteen eri klusterointimenetelmään. Tässä kappaleessa kuvaillaan molempien klusterointimenetelmien toteutus ja käyttö luokitteluun. Molemmissa tapauksissa etäisyysmittana käytetään euklidista etäisyyttä.

3.1 K:n keskipisteen menetelmä

Toteutettu algoritmi perustuu Han ja Kamberin [HKa01] kuvailemaan menetelmään suorittaa K:n keskipisteen klusterointi. Vaikka tehokkaampiakin algoritmeja on esitetty kirjallisuudessa (esim. [KMN02]), niin toteutuksessa päädyttiin silti yksinkertaisempaa menetelmään. Käytetty menetelmä on osoittautunut riittävän tehokkaaksi suoritettavaan tehtävään, jossa klusteroitavia vektoreita on alle 500 ja klustereita alle 20. Ohjelma käyttää enemmän aikaa itse äänitiedostojen lataamiseen kuin klusterointiin. Käytetty algoritmi eroaa Han ja Kamberin esittämästä algoritmista siten, että klustereiden alkukeskipisteiksi valitaan eri rumpuäänet. Toisin sanoen, ensimmäisen klusterin alkukeskipisteeksi valitaan klusterointi datan ensimmäinen bassorumpuääni, toisen klusterin keskipisteeksi ensimmäinen virveli, kolmannen ensimmäinen tom, jne.

Motivaatio K:n keskipisteen menetelmän käyttöön on piirrevektorien etsintä. Koska tiedetään, että datajoukko sisältää ääniä kahdeksasta eri instrumentista (rummuista ja pelleistä), niin kokeilemalla eri piirrevektoreita ja tarkastelemalla klusteroinnin tulosta voidaan päätellä testatun piirrevektorin hyvyys. Valitaan yksi piirrevektori, joka sisältää esimerkiksi tietoja äänen syttymisestä, ja suoritetaan klusterointi. Jos klusteroinnin tuloksena syntyy vain sellaisia klustereita, jotka sisältävät vain saman instrumentin ääniä, voidaan päätellä, että käytetty piirrevektori on hyvä. Jos kaikki klusterit sisältävät ääniä kaikista instrumenteista, niin voidaan päätellä, ettei piirrevektori kuvaa kovinkaan olennaisia asioita erottamisen kannalta.

Luokittelu K:n keskipisteen klusteroinnilla on toteutettu siten, että ensiksi klusteroidaan opetusdata. Kun klusterit on muodostettu, keskipisteille lasketaan tyyppi, eli arvo joka indikoi mitä rumpua klusteri edustaa. Klusterin tyypiksi valitaan eniten edustettu tyyppi ko. klusterista. Tämä saadaan käymällä läpi klusterin sisältämät äänet ja laskemalla kuinka monta kappaletta kutakin instrumenttia on edustettuna tässä klusterissa. Tyypin arvoksi tulee sitten edustetuimman instrumentin tyyppi. Tyyppi luokiteltavalle äänelle saadaan etsimällä opetusdatasta se ääni, joka on lähinnä (käyttäen samaa etäisyysmittaa kuin klusteroinnissa) luokiteltavaa ääntä. Opetusdatan äänestä katsotaan mihin klusteriin se kuuluu ja luokiteltava ääni luokitellaan klusterin tyyppiseksi. Syy siihen, miksi etäisyys lasketaan opetusdatan ääneen, eikä klustereiden keskipisteisiin on seuraava: Jos luokiteltava ääni on jonkin klusterin alueen sisäpuolella, niin lopputulos on sama kuin jos laskettaisiin etäisyys klusterien keskipisteisiin. Jos kuitenkin luokiteltava ääni on klusterin alueitten ulkopuolella, niin tällöin laskettu etäisyys on luokiteltavasta äänestä lähimpään klusterin rajaan eikä lähimpään keskipisteeseen.

3.2 Itseorganisoituva kartta

Toteutettu itseorganisoituva kartta perustuu Teuvo Kohosen esittämään rakenteeseen [Koh90]. Toteutuksessa on käytetty neliöllistä kaksiuloitteista topologiaa. Neuronin ympäristön säteen laskemiseksi käytetään kaavaa:

$$\theta(t) = Y - \sum_{t=0}^{T} (Y/T), \tag{1}$$

missä $\theta(t)$ on säteen pituus, Y on säteen alkuarvo ja T on opetusdatan iteraatioiden lukumäärä. Säteen pituus lyhenee siis lineaarisesti opetuksen edetessä kohti nollaa. Säteen alkuarvona käytetään arvoa, jolla neuronin ympäristön ala on vähän yli puolet kartan pinta-alasta.

Ympäristön opetukseen käytetään painoarvoa α, päivittäen sitä käyttäen kaavaa

$$\alpha(t) = 0.9(1 - t/K), \tag{2}$$

Rumpu	*Lkm*	*Luokitus %*
Bassorumpu	56	95 %
Virveli	81	94 %
Tom	12	100 %
Ride	17	76 %
Hihat	59	66 %
Hihat avoinna	46	59 %
Crash	12	100 %
Splash	5	100 %

Taulukko 3: Luokittelun tulos rumpukohtaisesti.

missä K on karkean opetusvaiheen iteraatioiden lukumäärä. Tämän jälkeen α:n arvo asetetaan 0.02:ksi. Hienosäätövaiheessa α:n arvoa vähennetään taas lineaarisesti siten, että opetuksen päätyttyä α:n arvo saavuttaa nollan.

Jokainen ympäristön neuroni opetetaan kaavalla:

$$m_i(t+1) = m_i(t) + \alpha(t)(x(t) - m_i(t)), \tag{3}$$

missä $m_i(t)$ on neuronin nykyinen painovektori ja $x(t)$ on syötetty opetusvektori.

Opetusdata syötetään kartalle satunnaisessa järjestyksessä useaan kertaa. Näin iteraatioita opetukselle tulee tarvittava määrä. Jokaiselle neuronille lasketaan tyyppi sen voittamien syötevektoreiden tyyppien perusteella. Jos neuroni on voittanut eniten bassorumpusyötteitä, niin neuroni tyypitetään bassorummuksi. Voittavalla neuronilla tarkoitetaan sitä kartan neuronia, joka on lähimpänä sen hetkistä syötevektoria.

Kuten K:n keskipisteenmenetelmässäkin, itseorganisoituvaa karttaa voidaan käyttää arvioimaan käytetyn piirrevektorin hyvyyttä. Jos karttaan muodostuu selviä yhtenäisiä alueita, jossa vierekkäiset neuronit ovat saman tyyppisiä, käytetty piirrevektori on hyvä. Jos kartta ei organisoidu, voidaan olettaa, ettei piirrevektori sisällä olennaista tietoa äänien erotteluun.

Ennen luokittelua itseorganisoituva kartta opetetaan opetusdatalla. Luokiteltavasta äänestä generoidaan piirrevektori, jonka jälkeen etsitään kartasta se neuroni, joka on lähinnä luokiteltavaa ääntä. Luokiteltava ääni luokitellaan tämän neuronin tyyppiseksi.

4 Tulokset

Menetelmässä käytetty koulutusdata sisälsi 400 rumpuääntä ja testidata 288 rumpuääntä. Koulutusdatassa jokaista rumpua edusti 50 eri ääntä. Testiaineistossa rumpujen määrät puolestaan vaihtelivat. Tämä johtui aineiston huonosta saatavuudesta.

Käytetty piirrevektori tuntui erottelevan rummut hyvin. Suurimpia virheitä ilmeni peltien erottelussa. Rumpujen osalta virheet olivat odotetun kaltaisia. Näitä olivat esimerkiksi lattia-tomin (matalataajuuksinen-tom) ja bassorummun sekoittaminen ja virvelin ja pikkutomin (korkeataajuuksinen-tom) sekoittaminen. Tälläiset virheet ovat hyvin tyypillisiä myös ihmiselle. Peltien osalta virheet olivat epäselvempiä. Taulussa 3 on esitelty rumpukohtaisesti luokittelun tulos. Luokitteluun on käytetty itseorganisoituvaa karttaa, joka luokitteli koko testiaineiston 81 %:n tarkkuudella.

Itseorganisoituvalla kartalla tulokset olivat parempia kuin k:n keskipisteen klusteroinnilla. Taulussa 4 on esitelty itseorganisoituvalla kartalla saatuja tuloksia. Taulussa 5 on esitetty

Leveys	*Korkeus*	*Iteraatioita*	*Luokitus %*	*Rumpu/Pelti -luokitus %*
5	5	8000	63	97
10	10	8000	77±2	97±1
10	10	16000	78±2	96±1
12	12	16000	79±2	96±2
15	15	16000	76±2	96±1
15	15	28000	79	97
20	20	28000	80	96±1
20	20	32000	78	96±1

Taulukko 4: Itseorganisoituvalla kartalla saadut tulokset.

Klustereiden lkm	*Luokitus %*	*Rumpu/Pelti -luokitus %*
8	60	97
10	67	98
11	65	97
12	72	98
13	66	98
14	65	98

Taulukko 5: K:n keskipisteen menetelmällä saadut tulokset.

k:n keskipisteen menetelmällä saatuja tuloksia. Vertaukseksi k:n lähimmän naapurin menetelmällä saatiin 87 %:n luokitus (99 % rummut/pelti luokitus), kun k:n arvona käytettiin arvoa 1 (eli lähin naapuri).

Vastaavissa toteutetuissa järjestelmissä on saavutettu noin 90 %:n luokittelu [HYG02].

5 Yhteenveto

Tässä paperissa esiteltiin toteutettu järjestelmä rumpuäänien luokitteluun. Luokitteluun käytettiin kahta eri menetelmää: k:n keskipisteen klusterointia ja itseorganisoituvaa karttaa. Luokittelun motivaationa oli kehittää järjestelmä, joka osaa nimetä syötteenä saadut rumpuäänet. Luokittelun toisena motiivina oli kehittää järjestelmä, jonka avulla voidaan arvioida rumpuäänestä generoitujen piirrevektorien hyvyyttä.

Toteutetulla järjestelmällä saatiin kohtalaisen hyviä tuloksia. Tuloksiin tulee suhtautua kuitenkin varauksella, sillä käsitelty aineisto on itse kerättyä ja vaihtelevan laatuista. K:n keskipisteen menetelmän huono luokittelumenestys osoittaa vielä, että käytetyssä piirrevektorissa on parantamisen varaa.

Järjestelmässä käytetyn piirrevektorin dimensio on 92, siis erittäin suuri. Jatkossa järjestelmää kehitetään etsimällä keinoja pienentää käytetyn piirrevektorin dimensiota.

Viitteet

[Ero01] Eronen, A., "Comparison of Features for Musical Instrument Recognition", *In Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, WASPAA 2001

[GoM95] Goto, M., Muraoka, Y., "A Real-time beat tracking system for audio signals", *In proceedings of International Computer Music Conference*, 171-174, 1995

[GPD00] Gouyon, F., Pachet, F., Delerue, O., "On the use of zero crossing rate for an Application of Classification of Percussive Sounds", *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00)*, Verona, Italy, December 7-9, 2000

[Gre77] Grey J.M., "Multidimensional perceptual scaling of musical timbres.", *Journal of Acoustical Society of America*, 61(5), 1977, s. 1270-1277.

[HKa01] Han, J., Kamber, M., "Data Mining: Concepts and Techniques", Morgan Kaufmann, 2001, s. 349-353

[HYG02] Herrera, P., Yeterian A., and Gouyon F.,"Automatic classification of drum sounds: a comparison of feature selection methods and classification techniques", 2002, citeseer.nj.nec.com/herrera02automatic.html

[KMN02] Kanungo, T., Mount, D. M., Netanyahu N. S., Piatko, C. D., Silverman, R., Wu, A. Y., "An Efficient k-Means Clustering Algorithm: Analysis and Implementation", *IEEE Transactions on pattern analysis and machine intelligence*, Vol. 24, No. 7, July 2002.

[Koh90] Kohonen Teuvo, The Self-Organizing Map. *Proceedings of the IEEE*, Vol. 78, No. 9, 1990

[LMe94] Laroche, J., Meillier, J.-L., "Multichannel excitation/filter modeling of percussive sounds with application to the piano.", *IEEE Transactions onSpeech and Audio Processing*, 2, (1994), 329-344

[Lyo82] Lyon, R. F., "a Computational model of filtering, detection and compression in the cochlea", *In Proceedings of the IEEE International conference on Acoustics, Speech, and Signal Processing*, Paris, May 1982.

[Med86] Meddis, R., "Simulation of Mechanical to neural transduction in the auditory receptor", *Journal of the Acoustical Society of America*, vol. 79, no.3, pp. 702-711, Mar. 1986.

[PRH92] Patterson, R. D.,Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., and Allerhand, M., "Complex sounds and auditory images", in Auditory Physiology and Perception, Y. Cazals, L. Demany, and K. Horner, Editors, vol. 83 of Advances in the Biosciences, pp. 429-445. PergamonPress, Oxford, 1992, *Proceedings of the 9th International Symposium on Hearing* held in Carcens, France, June 1991.

[Sla88] Malcolm Slaney, "An efficient implementation of the Patterson-Holdsworth auditory filter bank", *Apple Computer Technical Report 35*, Apple Computer, 1988, http://www.slaney.org/malcolm/pubs.html.