

PARAMETRISOITU TILÄÄNENTOISTO JA -SYNTEESI VIRTUAALIMAAILMOISSA

Tapani Pihlajamäki¹

¹ Aalto-yliopiston sähkötekniikan ja elektroniikan korkeakoulu
Signaalinkäsittelyn ja akustiikan laitos
Otakaari 5, 02150 Espoo
tapani.pihlajamaki@aalto.fi

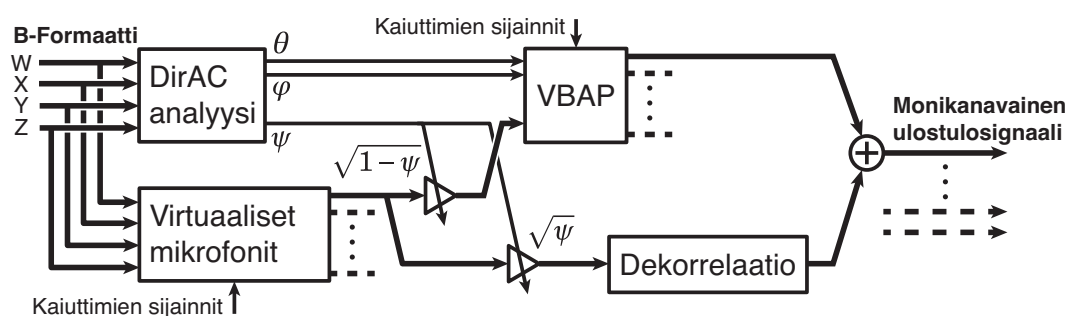
Tiivistelmä

Perinteisesti äänentoisto ja äänisynteesi virtuaalimaailmoissa, erityisesti videopeleissä, perustuu menetelmiin, joissa ääni paikannetaan pistemäisinä lähteinä kuulijan ympärille ja sitä käsitellään erilaisilla audioefekteillä aikatasossa. Tämä perusmalli on ollut käytössä pitkään ja edistystä on tapahtunut lähinnä audioefektien ja tilamallien saralla. Tässä artikkelissa esitellään, miten tiläänentoistoon tarkoitettua Directional Audio Coding -menetelmää voidaan soveltaa korvaamaan virtuaalimaailmojen äänijärjestelmien perinteinen rakenne. Tämä parametrinen ratkaisu perustuu ihmiskuulon toimintaan ja sen ansiosta voidaan äänisuunnittelijalle tarjota vanhojen työkalujen lisäksi uudenlaisia mahdollisuuksia, joilla voidaan syntesoida helposti ja luontevasti esimerkiksi äänilähteille havaittava leveys. Lisäksi tämä menetelmä on laskennallisesti tehokas tapa luoda monimutkaisia äänimaisemia.

1 JOHDANTO

Virtuaalimaailma on nimensä mukaisesti virtuaalinen maailma, joka yleensä pyrkii tarjoamaan käyttäjälleen immersiiivisen kokemuksen läsnäolosta ja vuorovaikutuksesta keinotekoisessa ympäristössä. Monet nykyaikaiset videopelit ovatkin hyvä esimerkki tällaisista maailmoista. Oleellinen osa käyttäjän immersiota on laadukas ja uskottava äänimaailma, joka vastaa virtuaalimaailmassa tapahtuvia muuten havaittavia ilmiöitä. Suuren yleisön kannalta merkittävin kehitys virtuaalimaailmoissa alkoi 90-luvulla, kun videopelien suuntaus lähti laskentatehojen kasvaessa kohti immersiiivisiä kolmiulotteisia maailmoja. Samalla pelien ääniteknologia kehittyi, ja syntyi malli, jossa ääntä panoroidaan pistelähteinä, ja äänimaisemaan sovelletaan erilaisia tilamalleja ja audioefektejä [1]. Nykyään kaupalliset järjestelmät (esim. FMOD ja Audiokinetic Wwise) on suunniteltu saman perusasetelman päälle niin, että äänisuunnittelijan työ muodostuu helpommaksi ja muistuttaa yhä enemmän esimerkiksi elokuvamiksaajan työtä.

Tästä mallista voidaan kuitenkin siirtyä teknologisesti eteenpäin soveltamalla uusia äänentoiston ja -synteesin menetelmiä. Tuloksena äänisuunnittelijalle saadaan työn avuksi esimerkiksi proseduraalista ääntä [2], tarkempia tilamalleja [3, 4, 5] ja uusia rakenteita äänenkäsittelyyn [6, 7, 8]. Tässä artikkelissa esilläänkin uusi tapa lähestyä virtuaalimaailmojen äänentoistoa ja -synteesiä käyttäen parametrista tiläänentoistoa lähtökohtana.



Kuva 1: DirAC-menetelmän lohkokkaavio.

2 DIRECTIONAL AUDIO CODING

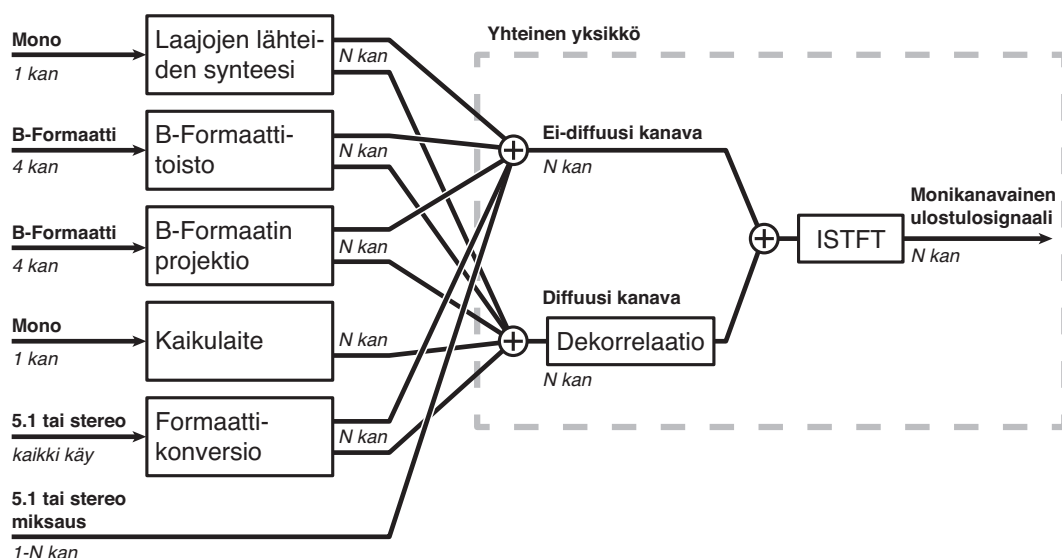
Directional Audio Coding (DirAC) on parametrinen esitysmuoto ja menetelmä tiläänen tallennukseen ja toistoon [9]. Se perustuu psykoakustiseen oletukseen, että ihminen pysyy yhdellä ajanhetkellä havaitsemaan vain yhden tulosuunnan ja yhden binauraalisen koherenssivihjeen äänelle jokaista kriittistä kaistaa kohden. Tämän oletuksen pohjalta on muodostettu parametrinen esitysmuoto tiläänellem, jossa ääni tallennetaan joko B-formaattisignaalin tai monosignaalin ja sen lisäksi talletetaan jokaiselle taajuuskaistalle äänen tulosuunta- ja diffuusisuusparametri. Tähän esitysmuotoon päästään normaalista B-formaattitallenteesta (tai muista tallennusformaateista) soveltamalla DirAC-analyysiä ja vastaavasti esitysmuoto voidaan toistaa käyttämällä DirAC-synteesiä. Tämän järjestelmän rakenne on esitetty kuvassa 1. Huomattavaa on, että tällä menetelmällä tilääntä voidaan myös siirtää merkittävästi pienemmällä tiedonsiirtokapasiteetilla kuin yleensä vaadittaisiin monikanavaisessa toistossa erillisten kaiutinsignaalien siirtämiseen.

2.1 DirAC-analyysi

DirAC-analyysi perustuu B-formaattisignaaliin tallennetun äänikentän energettiseen analyysiin. B-formaattisignaali itsessään muodostuu neljästä mikrofonisignaalista, joista w on pallokuviomikrofonilla äänitetty signaali, ja x , y ja z ovat suorassa kulmassa kolmeen eri suuntaan osoittavien dipolikuvioiden mikrofoniin signaalit. Näiden signaalien avulla voidaan estimoida ajasta ja taajuudesta riippuvat paine- ja hiukkaunopeussignaalit (p ja \vec{v}), joista vastaavasti pystytään estimoimaan äänikentän intensiteetti ja energia (\vec{I} ja E) tallennuspisteessä. Tämän jälkeen on mahdollista estimoida äänen tulosuunta $[\theta \varphi]$ ja diffuusisuus ψ . Tämä johto on esitetty Pulkin artikkelissa [9].

2.2 DirAC-synteesi

Tallennettu B-formaattisignaali (tai monosignaali) toistetaan analysoitujen parametrien avulla DirAC-synteesissä. Ensimmäiseksi B-formaattisignaalista muodostetaan virtuaalimikrofonit osoittamaan kaiuttimien suuntaan käytetyssä kaiutinjärjestelmässä niin, että kuuntelupiste vastaa mikrofoniin sijaintia alkuperäisessä tallenteessa. Tämän jälkeen signaalin energia jaetaan kahteen osaan riippuen analysoidusta diffuusisuusarvosta. Diffuusiksi määritetyn signaaliosan kanavien välinen korrelaatio poistetaan mahdollisim-



Kuva 2: VW-DirACin komponentit lohkokkaaviossa. Yhteiseen toistoyksikköön voidaan yhdistellä erilaisia esikäsittely-yksiköitä haluttu määrä.

man hyvin kuitenkin muuttamatta äänenväriä ja nämä dekorreloidut kaiutinsignaalit toistetaan sellaisenaan. Ei-diffuusiksi määritetty osuus toistetaan suuntaavana äänenä käyttäen vector-base amplitude panning (VBAP) -menetelmää [10] ohjaamaan kaiutinkanavien suhteellisia voimakkuuksia. Tuloksena siis kuuntelijalle toistetaan yksi suuntaava ääni ja diffuusi ympäröivä ääni jokaista taajuuskaistaa kohden, ja lopputuloksena oleva kuuntelijan kuulohavainto pyrkii mahdollisimman hyvin vastaamaan alkuperäisessä tallennustilanteessa saatavaa kuulohavaintoa.

3 PARAMETRINEN TILAÄÄNEN SYNTEESI

Normaali DirAC-menetelmä soveltuu hyvin tilanteisiin, joissa oikean maailman äänikenttä voidaan tallentaa ja toistaa. Virtuaalimaailmoissa valmista äänikenttää ei kuitenkaan ole vaan se on tarkoitus luoda synteettisesti, joten normaali DirAC soveltuu lähinnä ennaltatallennettujen äänimaisemien toistamiseen. Sen sijaan samojen lähtöoletuksien pohjalta voidaan luoda menetelmiä, joilla synteettisen äänimaiseman koostaminen on äänisuunnittelijalle luontevaa. Tätä menetelmien yhdistelmää kutsutaan tässä artikkelissa nimellä Virtual-world Directional Audio Coding (VW-DirAC) [8, 11] ja se tarjoaa äänisuunnittelijan käyttöön erilaisia uusia työkaluja. Yksinkertaisimmillaan nämä työkalut mahdollistavat jo olemassa olevat perustehtävät kuten pistelähteiden sijoittamisen äänimaisemaan ja tilavaikutelman luomisen kaikulaitteiden avulla. Kuuloon perustuvan parametrinen lähestymistavan ansiosta voidaan kuitenkin myös syntetisoida monimutkaisempia äänikentän ominaisuuksia kuten äänilähteen leveys ja diffuusisuus. Lisäksi jo olemassa olevia ominaisuuksia, kuten suuntajakaumaa, voidaan muokata tarkoitukseen sopivilla efekteillä. Kokonaisuutena VW-DirAC voidaan ajatella virtuaalimaailmojen ja pelien äänimoottorin korvaajana. Kuvassa 2 onkin nähtävissä eräs rakenne tällaiselle äänimoottorille käyttäen VW-DirAC:n tarjoamia menetelmiä ja seuraavissa kappaleissa käsitellään osaa kuvan esikäsittely-yksiköistä.

3.1 Havainnollisesti leveiden lähteiden luominen

Äänilähteen havaittavan koon muokkaaminen on uusista parametrinen synteessin tarjoamista menetelmistä rakenteeltaan yksinkertaisimpia, mutta toiminnaltaan äänisuunnittelijalle hyödyllisimpiä työkaluja. Käytännössä tämä menetelmä perustuu sisääntulevan monosignaalin jakamiseen aikataajuuskomponentteihin (esim. lyhytaikaisella Fourier-muunnoksella), jotka sirotellaan virtuaaliavaruudessa halutulle alueelle satunnaisesti tasajakaumaa käyttäen. Tuloksena syntyy ääni, joka havaitaan selkeästi leveänä äänilähteenä. Samalla äänenväri myös muuttuu, mutta muutosta ei havaita välttämättä laatua heikentävänä tekijänä [8]. Työkaluna tämä menetelmä tarjoaa yksinkertaisen parametrin äänilähteen koolle, jota äänisuunnittelija voi muokata tai äänimoottori voi ohjata automaattisesti virtuaalimaailman komponenttien mukaisesti.

3.2 B-formaattiäänitteen projektio

B-formaattiäänitteen projektio on jatkosovellus leveiden lähteiden synteessille ja normaalille DirAC:n B-formaattitoistolle. Normaalisissa B-formaattitoistossa kuuntelija on rajoitettu tallennetun äänimaiseman suhteen samaan paikkaan kuin missä mikrofoni sijaitsee äänitystä tehtäessä. Virtuaalimaailman itseisarvona on kuitenkin yleensä se, että käyttäjä pystyy liikkumaan maailmassa. Tällöin kiinteä B-formaattitoisto ei välttämättä sovellu hyvin tallennetun äänimaiseman toistoon.

Tämä ongelma voidaan ratkaista yhdistämällä DirAC-analyysi leveän lähteen synteisiin [12]. DirAC-analyysistä saadaan tieto äänilähteiden suunnista aikataajuusriippuvasti. Sirottelemalla w -signaalin aikataajuuskomponentit analysoituihin suuntiin, saadaan ääni levitettyä havainnollisesti haluttuihin suuntiin. Äänen etäisyys pitää kuitenkin rajoittaa jollekin etäisyydelle, joten sirottelu suoritetaan projisoimalla aikataajuuskomponentit piste-lähteiksi mielivaltaisen konveksin avaruuskappaleen (esim. pallo tai sylinteri) pinnalle. Tätä avaruuskappaletta voidaan sitten muuntaa ja siirtää suhteessa käyttäjän hahmoon, jolloin saadaan tuotettua havainto liikkeestä suhteessa tallennettuun äänimaisemaan.

Käytännössä tätä menetelmää voidaan soveltaa helposti tuottamaan valmiita äänimaisemia tiloihin ja alueille virtuaalimaailmoissa niin, että äänessä säilyy luonnollinen interaktiivisuus.

3.3 Parametriset tilääniefektit

Esitysmuotona parametrinen tilääni mahdollistaa myös erilaisia ääniefektejä, joilla tallennettua tilääntä voidaan muokata käsittelemällä sen parametreja. DirAC-pohjaisessa parametriseissa tiläänessä nämä efektit voidaan jakaa karkeasti kahteen luokkaan, joista ensimmäisessä muokataan äänen suuntainformaatiota ja toisessa äänen kaiku- ja ambienssiominaisuuksia. Seuraavaksi esitellään muutamia näistä efekteistä, jotka on aiemmin julkaistu Digital Audio Effects -konferenssissa [13].

Yksinkertaisimmillaan analysoituja äänen tulosuuntia voidaan pyörittää kuuntelupisteen ympärillä ja samalla on myös mahdollista supistaa tai laajentaa käytettyä aluetta (eli kompressio ja ekspansio). Käytännössä nämä toiminnot ovat hyvin yksinkertaisia ja

voidaan kuvata tulosuunnasta riippuvalla siirtofunktiolla. Samaan kategoriaan voidaan myös laskea suunnan mukaan äänien suodattaminen, joka siis mahdollistaa periaatteessa mielivaltaisen suuntakuvion luomisen. Astetta monimutkaisempi muunnos saadaan, kun siirretään kuuntelupistettä eli toteutetaan kameramainen zoom-efekti äänelle. Kappaleessa 3.2 oleva B-formaattiaänitteen projektio on tämän algoritmin yleistäpaus, mutta tiukemmin rajatussa efektiikäytössä voidaan mukaan tuottaa enemmän efektimäisiä piirteitä, jotka tehostavat haluttua vaikutelmaa. Neljäs tapa muuntaa suuntainformaatio on yhden B-formaattiaänitteen suuntajakauman käyttämien toisen B-formaattiaänitteen suuntajakaumana. Tätä kutsutaan spatiaaliseksi modulaatioksi. Käytännössä tämä menetelmä on tarkoitettu puhtaasti efektiikäyttöön ja mahdollistaa myös suuntainformaatioiden yhdistelyn.

Vastaavasti kaiku- ja ambienssiominaisuuksia voidaan muokata diffuusisuusparametrin avulla. Yksinkertainen tapa on antaa äänisuunnittelijalle käyttöön taajuusriippuva käyrä (kuten graafisessa ekvalisoinnissa), jonka avulla on mahdollistaa säätää tiläänitallenteessa suoran äänen ja jälkikaiunnan suhdetta (direct-to-reverberant ratio eli DRR). Tällä efektilä on helppoa esimerkiksi taajuuskohtaisesti korostaa suoraa ääntä. Toinen efekti voidaan tuottaa niin, että pyritään vaimentamaan suoraa ääntä ja siten toistamaan äänitteen ambienssi mahdollisimman hyvin. Käytännössä tämä toteutetaan niin, että luodaan virtuaalinen mikrofoni osoittamaan pois vahvimasta analysoidusta tulosuunnasta ja diffuusisuusparametrilla dynaamisesti muokataan tämän mikrofonin suuntakuviota kardiodin ja pallokuvion välillä niin, että pieni diffuusisuusarvo tuottaa suuntaavamman virtuaalimikrofonin.

4 DISKUSSIO JA YHTEENVETO

Tässä artikkelissa esitellyt menetelmät tarjoavat äänisuunnittelijoille uudenlaisia työkaluja äänimaisemien luomiseen. Vaikka nämä menetelmät pohjautuvatkin monelta osin Directional Audio Coding -tekniikkaan, niin keskeinen yhtymäkohta on kuitenkin vain se, että tarpeellista on tuottaa riittävän uskottava havainto halutusta äänimaisemasta ihmiselle. Esiteltyjen menetelmien etuna on se, että äänisuunnittelija voi muokata äänimaisemaa sellaisten parametrien avulla (esim. suunta, leveys ja ympäröivyyys), jotka ovat usein hyvin luonnollisia käsittää, ja menetelmät perustuvat juuri näiden parametrien ympärille. Samalla voidaan myös helposti yksinkertaistaa äänimaisemia tarpeen vaatiessa ja siten saavutetaan laskennallisesti tehokas tapa toistaa ja syntetisoida monimutkaisia äänimaisemia.

5 KIITOKSET

Kiitokset Mikko-Ville Laitiselle, Ville Pulkille ja Archontis Politiksellle avusta tähän artikkeliin johtaneessa tutkimuksessa.

The research leading to these results has received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement n° [240453].

VIITTEET

- [1] Andy Farnell. Sound synthesis for games. In *Sounding Out 2006*, 2006.
- [2] Charles Verron, Mitsuko Aramaki, Richard Kronland-Martinet, and Grégory Pallo-ne. A 3-D immersive synthesizer for environmental sounds. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(6):1550–1561, August 2010.
- [3] Nicolas Tsingos, Emmanuel Gallo, and George Drettakis. Perceptual Audio Rende-ring of Complex Virtual Environments. In *Proceedings of ACM SIGGRAPH 2004*, pages 249–258, 2004.
- [4] Nicolas Tsingos. Pre-computing geometry-based reverberation effects for games. In *Proceedings of AES 35th International Conference*, 2009.
- [5] C. Schissler and D. Manocha. GSound: Interactive Sound Propagation for Games. In *Proceedings of 41st International Conference of AES: Audio for Games*, London, UK, February 2011.
- [6] Richard Furse. Building an OpenAL Implementation Using Ambisonics. In *Proceedings of AES 35th International Conference: Audio For Games*, London, UK, February 2009.
- [7] Leonid Terentiev, Cornelia Falch, Oliver Hellmuth, Johannes Hilpert, Werner Oo-men, Jonas Engdegård, and Harald Mundt. Saoc for gaming - the upcoming mpeg standard on parametric object based audio coding. In *Proceedings of AES 35th International Conference: Audio For Games*, London, UK, February 2009.
- [8] Mikko-Ville Laitinen, Tapani Pihlajamäki, Cumhur Erkut, and Ville Pulkki. Pa-rametric time-frequency representation of spatial sound in virtual worlds. *ACM Transactions on Applied Perception*, 9(2), 2012.
- [9] Ville Pulkki. Spatial sound reproduction with directional audio coding. *Journal of Audio Engineering Society*, 55(6):503–516, June 2007.
- [10] Ville Pulkki. Virtual sound source positioning using vector base amplitude panning. *Journal of Audio Engineering Society*, 45(6):456–466, June 1997.
- [11] Tapani Pihlajamäki, Mikko-Ville Laitinen, and Ville Pulkki. Modular Architecture for Virtual-World Parametric Spatial Audio Synthesis. In *Proceedings of 49th International Conference of AES: Audio for Games*, London, UK, February 2013.
- [12] Tapani Pihlajamäki and Ville Pulkki. Projecting Simulated or Recorded Spatial Sound onto 3D-Surfaces. In *Proceedings of 45th International Conference of AES: Applications of Time-Frequency Processing in Audio*, Helsinki, Finland, March 2012.
- [13] Archontis Politis, Tapani Pihlajamäki, and Ville Pulkki. Parametric Spatial Audio Effects. In *Proceedings of 15th International Conference on Digital Audio Effects (DAFx-12)*, York, UK, September 2012.