

# ERILLISMIKROFONIÄÄNITYSTEN KÄYTTÖ PARAMETRISISSA TILÄÄNEN KOODAAMISESSA

**Mikko-Ville Laitinen**<sup>1</sup>

<sup>1</sup> Aalto-yliopiston sähkötekniikan ja elektroniikan korkeakoulu  
Signaalinkäsittelyn ja akustiikan laitos  
Otakaari 5, 02150 Espoo  
mikko-ville.laitinen@aalto.fi

## Tiivistelmä

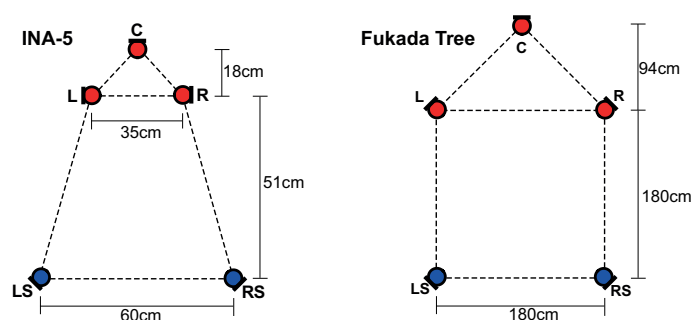
Erillismikrofonitekniikoita käytetään yleisesti monikanavaäänityksissä. Näissä tekniikoissa mikrofonien välinen koherenssi on pieni niiden etäisyyksistä johtuen, mikä johtaa miellyttävään ”ympäröivään” tilantuntuun. Valitettavasti erillismikrofonitekniikat kuitenkin kärsivät epätarkasta äänilähteiden lokalisaatiosta. Tässä artikkelissa esitetään Directional Audio Coding -tekniikkaan (DirAC) pohjautuva menetelmä erillismikrofoniiäänitysten parametriseen prosessointiin. Menetelmässä analysoidaan taajuuskaistoittain äänilähteiden suunta sekä diffuusisuus, minkä avulla toisto pyritään tekemään optimaalisesti. Menetelmä parantaa havaittua laatua perinteisiin erillismikrofonitekniikoihin verrattuna tarjoamalla tarkempaa ja stabiilimpaa äänilähteiden lokalisaatiota, kuitenkin siten, että hyvä tilantuntu säilyy. Erillismikrofoniiäänitysten alhaisen mikrofonien välisen koherenssin takia voidaan parantaa myös perinteisen DirAC-tekniikan laatua. Artikkelissa esitetään prosessointi muutamalle tunnetulle erillismikrofonisijoittelulle, sekä menetelmää varten optimoidulle mikrofonisijoittelulle.

## 1 JOHDANTO

Tiläänen tallentamiseen monikanavaäänentoistoa varten on useita tunnettuja tekniikoita [1]. Yleensä nämä tekniikat jaetaan karkeasti kahteen ryhmään: koinsidenssi- ja erillismikrofonitekniikoihin (coincident and spaced-microphone techniques).

Koinsidenssitekniikoissa käytetään suuntaavia mikrofoneja, joiden oletetaan olevan käytännössä yhdessä pisteessä. Kaiutinsignaalit saadaan mikrofonisignaalien painotettuna tai suodatettuna summana, kuten käyttämällä tunnettua Ambisonics-tekniikkaa [2]. Yleisesti käytössä olevien mikrofonien leveiden suuntakuvioiden takia kaiutinsignaalit ovat kuitenkin verrattain koherentteja, mikä aiheuttaa äänen sävyyn ja tilantuntuun liittyviä artefakteja. Tätä ongelmaa ratkaisemaan kehitettiin Directional Audio Coding -tekniikka (DirAC) [3], missä äänikentästä pyritään analysoimaan psykoakustisesti merkityksellistä suuntainformaatiota, jonka avulla toisto pyritään tekemään siten, että toistettu äänikenttä havaitaan samalla tavalla kuin alkuperäinen äänikenttä.

Vastaavasti erillismikrofonitekniikoissa mikrofonien väliset etäisyydet ovat verrattain suuria, kymmenistä senttimetreistä jopa useisiin metreihin. Näissä tekniikoissa mikrofonisignaalit voidaan syöttää suoraan kaiuttimiin. Saapuva tasoalto tallentuu käytännössä



Kuva 1: Erillismikrofonitekniikoiden esimerkkiaseteluja. Mikrofonit ovat kardioidoja.

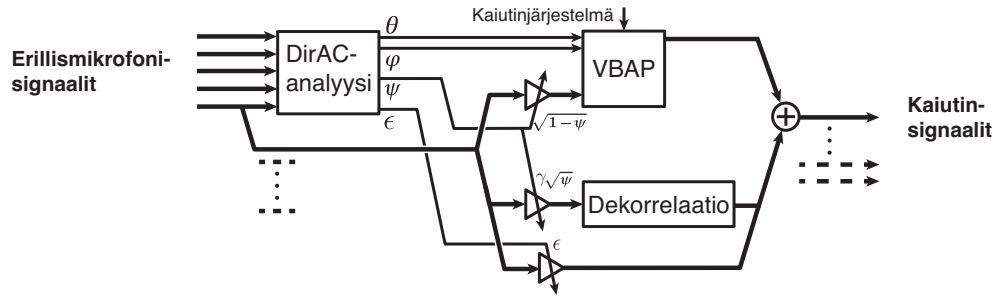
kaikkiin mikrofoneihin, ja suuntahavainto perustuu viiveisiin ja tasoeroihin kaiutinkana-  
vien välillä. Valitettavasti havaittu suunta on kuitenkin usein epämääräinen ja virheellinen  
[4]. Toisaalta erillismikrofonitekniikoissa mikrofoni-  
välinen koherenssi on verrattain  
matala, mikä mahdollistaa miellyttävän äänen sävyn sekä ympäröivän tilantun-  
nun.

Tässä artikkelissa esitetään menetelmä [5, 6, 7], jonka avulla erillismikrofonii-  
äänityksiä  
voidaan käsitellä parametrisesti käyttämällä DirAC-tekniikasta tunnettuja menetelmiä.  
Tarkoituksena on parantaa suuntahavainnon laatua, mutta kuitenkin pyrkiä säilyttämään  
erillismikrofonii-  
äänitysten hyvä tilantuntu. Lisäksi perinteisen DirAC-tekniikan laatua  
pyritetään parantamaan käyttämällä hyödyksi mikrofoni-  
välistä pientä koherenssia.

## 2 TAUSTAA

DirAC-tekniikan [3] perusajatus on, että äänikenttää ei tarvitse toistaa fyysisesti oi-  
kein. Riittää, että ihmisen havainto on oikea. Lisäksi oletetaan, että yhdellä ajanhetkellä  
yhdellä kuulon kriittisellä kaistalla ihminen voi havaita vain yhden suunnan sekä kor-  
vien välisen korrelaation. Toistamalla nämä suuntavihjeet ihmisen kuuloa vastaavalla  
aika- ja taajuusresoluutiolla, tila-  
äänien havainnon tulisi vastata alkuperäistä äänikenttää. Voidakseen toistaa nämä, äänikentästä tulee analysoida vastaavia ominaisuuksia, jotka  
ovat DirAC-tekniikassa äänen tulosuunta ja diffuusisuus. Ne analysoidaan äänitetyistä  
B-formaattisignaaleista aikariippuvina taajuuskaistoittain. Diffuusisuus kuvaa sitä, ole-  
tetaanko äänikentän koostuvan pelkästään yhdestä tasoallostasta, pelkästään diffuusista  
kentästä vai jotain siltä väliltä. Synteesissä DirAC-tekniikassa pyritään kontrolloimaan  
ihmisen suunnan ja koherenssin havaintoa analyysidatan perusteella. Ääni jaetaan aluksi  
ei-diffuusiin ja diffuusiin ääneen. Ei-diffuusi ääni pyritään toistamaan amplitudipanoro-  
inin avulla halutusta suunnasta. Vastaavasti diffuusi ääni pyritään toistamaan epäkohe-  
renttina dekorreloinnin avulla. Näitä kahta osaa painotetaan analysoidun diffuusisuuden  
perusteella, minkä tuloksena sekä havaitun suunnan että korvien välisen koherenssin, ja  
siten myös tila-  
äänien havainnon, tulisi vastata alkuperäistä äänikenttää.

Perinteisissä erillismikrofonitekniikoissa parametrin analyysin sijaan äänikentän ha-  
vainto pyritään toistamaan mikrofoni-  
sijoittelun avulla sekä valitsemalla sopivat suun-  
takuviot. Kuvassa 1 on nähtävissä kaksi tunnettua erillismikrofonitekniikkaa [8]: INA-5  
(Ideale Nieren-Anordnung) ja Fukada Tree. Kummassakin tekniikassa voidaan nähdä sa-  
ma periaate: edessä on kolme mikrofonia, joiden avulla pyritään toistamaan äänilähteiden  
suunta, ja takana on kaksi mikrofonia, joilla pyritään toistamaan tilantuntua.



Kuva 2: Erillismikrofoniäänitusten parametrinen käsittelyn lohkokkaavio.

### 3 ERILLISMIKROFONIÄÄNITYSTEN PARAMETRINEN KÄSITTELY

Sekä DirAC-tekniikalla että erillismikrofonitekniikoilla voidaan toistaa tilääntä hyvällä laadulla. Molemmissa on kuitenkin ongelmansa. DirAC-tekniikassa eräs ongelmista on dekorrelaatio, joka tietyillä signaaleilla saattaa huonontaa laatua. Vastaavasti erillismikrofonitekniikoissa suuntien havainnointi on melko epämääräistä. Tässä artikkelissa esitetyllä menetelmällä pyritään kiertämään kummankin lähestymistavan ongelmat ja yhdistämään hyvät puolet. Prosessointi tehdään kahdessa vaiheessa: analyysi ja synteesi.

Äänikentästä pyritään analysoimaan perinteisen DirAC-tekniikan tapaan suunta ja diffuusisuus taajuuskaistoittain [6]. Oletetaan, että on kaksi suuntamikrofonia, kuten kardiodia, jotka ovat symmetrisessä kulmassa x-akseliin nähden. Laskemalla näiden kahden signaalin magnitudispektrien erotus, saadaan tulokulmaan verrannollinen muuttuja

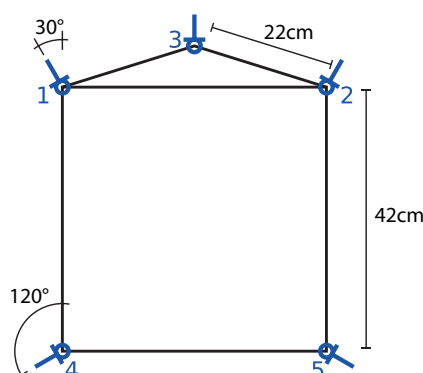
$$I_x(f) = |S_1(f)| - |S_2(f)| = |A(f)|(1 - \alpha)2 \sin \theta_0 \cos \theta, \quad (1)$$

missä  $|A(f)|$  on saapuvan tasoallon magnitudispektri,  $\alpha$  mikrofonin suuntaavuusparametri,  $\theta_0$  mikrofonien välinen kulma ja  $\theta$  saapuvan tasoallon suunta. Asettamalla vastaavasti kaksi mikrofonia y-akselin suhteen symmetrisesti, voidaan luoda suuntavektori  $\mathbf{I}_{\text{mag}} = [I_x \ I_y]$  ja laskea tulokulma  $\theta = \tan^{-1}(I_y/I_x)$ . Diffuusisuus voidaan laskea suuntavektorin ajassa tapahtuvaan vaihteluun perustuen

$$\psi = \sqrt{1 - \frac{\|\mathbf{E}\{\mathbf{I}_{\text{mag}}\}\|}{\mathbf{E}\{\|\mathbf{I}_{\text{mag}}\|\}}}, \quad (2)$$

missä  $\mathbf{E}$  tarkoittaa odotusarvoa, mikä lasketaan käytännössä aikakeskiarvoistuksella.

Synteesin lohkokkaavio on nähtävissä kuvassa 2. Perinteisen DirAC-tekniikan tapaan mikrofonisignaalit jaetaan aluksi ei-diffuusiin ja diffuusiin ääneen. Siitä poiketen kuitenkin vain osa diffuusista äänestä dekorreloidaan. Syy tähän on se, että siinä missä perinteisessä DirAC-tekniikassa käytetyt B-formaattisignaalit ovat melko koherenteja, erillismikrofonisignaalien välinen koherenssi on keski- ja korkeilla taajuuksilla varsin pieni. Tämän takia niitä ei suurelta osin tarvitse dekorreloida, millä vältetään dekorreloinista johtuvat artefaktit. Jako dekorreloituun ja dekorreloimattomaan diffuusiin ääneen tehdään mikrofonien välillä taajuuskaistoittain estimoidun koherenssin  $\gamma$  perusteella. Estimointi voidaan tehdä esimerkiksi olettamalla diffuusi kenttä [9]. Ei-dekorreloidun diffuusiin äänen vahvistus  $\epsilon$  määritetään lopuksi siten, että kokonaisenergia säilyy.



Kuva 3: Esitettyä menetelmää varten optimoitu kardioideja käyttävä mikrofoniasettelu.

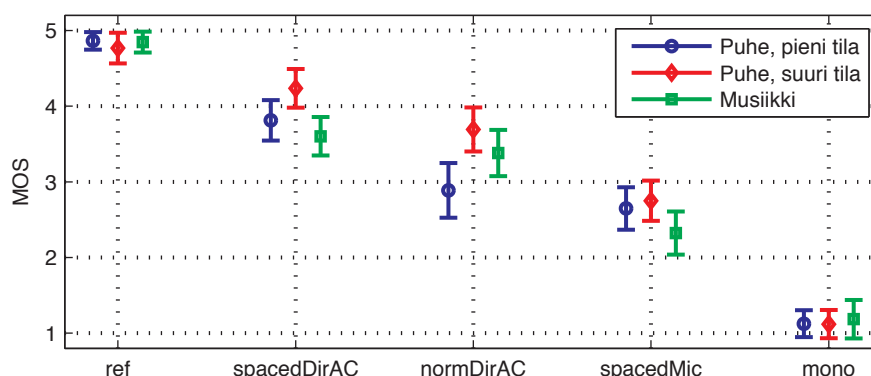
## 4 KÄYTÄNNÖN TOTEUTUKSIA

Kappaleessa 3 esitettyä menetelmää voi käyttää olemassa olevien erillismikrofoniasettelujen kanssa, mutta mikrofoniasettelu voidaan myös yrittää optimoida menetelmää varten. Analyysin puolesta a) mikrofonien väliset kulmat pitäisivät olla sopivat analyysia varten ja b) mikrofonien välinen etäisyys pitäisi olla minimaalinen [6]. Synteesin puolesta a) mikrofonien tulisi osoittaa kaiuttimia ja vierekkäisten mikrofonien välinen aikaero tulisi minimoida [6] ja b) mikrofonien välinen etäisyys pitäisi olla mahdollisimman suuri, mielellään vähintään 40 cm [5]. Vaatimukset ovat ilmeisesti ristiriitaiset, ja sopiva ratkaisu on aina kompromissi. Esimerkkiratkaisu ongelmaan on nähtävissä kuvassa 3. Siinä etumikrofonit on pyritty asettelemaan siten, että edessä olevien äänilähteiden suunta voitaisiin toistaa mahdollisimman tarkasti. Takamikrofoneilla yleensä toistetaan lähinnä kaiuntaa ja ambienssia, joten niiden osalta tarkka lokalisatio ei ole niin tärkeää. Siksi niiden asettelu on valittu siten, että suunta-analyysi edessä olevia lähteitä varten voidaan tehdä tarkasti, ja että signaalien välillä on sopiva koherenssi.

Kuvassa 1 esitettyjen mikrofoniasettelujen tapauksessa analyysi tehdään käyttämällä vain edessä olevia mikrofoneja (L, C ja R), jotka ovat riittävän lähellä toisiaan suunta-analyysia varten. Näissä asetteluissa mikrofoneja ei ole sopivissa kulmissa, jotta kappaleessa 3 esitettyä analyysia voitaisiin käyttää suoraan. Analyysia hieman muokkaamalla voidaan kuitenkin päästä riittävän tarkkaan tulokseen myös näillä asetteluilla [7].

## 5 KUUNTELUKOKEET

Tässä artikkelissa esitetyn menetelmän toiston laatua evaluoitiin formaaleilla kuuntelukokeilla [6]. Kuuntelukoe toteutettiin hyvin akustoidussa kuunteluhuoneessa. Toistojärjestelmässä oli yhdeksän kaiutinta 0 asteen korkeudessa suunnissa  $0^\circ$ ,  $\pm 30^\circ$ ,  $\pm 70^\circ$ ,  $\pm 110^\circ$  ja  $\pm 160^\circ$ , sekä neljä kaiutinta 45 asteen korkeudessa suunnissa  $\pm 45^\circ$  ja  $\pm 135^\circ$ . Kokeeseen osallistui 18 kuuntelijaa. Testimateriaalina käytettiin kolmea ääninäytettä: samanaikaista puhetta eri suunnista sekä pienessä että suuressa huoneessa ja musiikkia keskikokoisessa huoneessa. Ääninäytteet valittiin siten, että ne tiedettiin ennestään haastaviksi DirAC-tekniikoille. Testissä käytetyt huoneet simuloitiin kovalähdemenetelmällä. Tuloksena saatiin signaalit kaiutinjärjestelmässä oleville kaiuttimille siten, että suorat äänet ja jokai-



Kuva 4: Kuuntelukokeissa havaittu yleisen laadun heikkeneminen verrattuna referenssiin. Keskiarvot ja 95%:n luottamusvälit ovat nähtävillä kuvassa.

nen heijastus oli kvantisoitu lähimpään kaiuttimeen. Näiden referenssisignaalien avulla simuloitiin myös B-formaatti- ja erillismikrofonisignaalit.

Testissä oli tutkittava viisi eri menetelmää. Piilotettu referenssi (*ref*) oli identtinen luodun 13-kanavaisen referenssin kanssa. Monoversio (*mono*) sisälsi referenssin kanavien summan toistettuna edessä olevalla kaiuttimella. Muissa menetelmissä toisto tehtiin 5.1:n mukaisilla viidellä kaiuttimella ( $0^\circ$ ,  $\pm 30^\circ$ ,  $\pm 110^\circ$ ). Kuvassa 3 esitetyn asettelun mukaiset signaalit toistettiin suoraan ilman käsittelyä menetelmässä *spacedMic*. Samat signaalit toistettiin tässä artikkelissa esitetyllä parametrillisella tekniikalla menetelmässä *spacedDirAC*. B-formaattisignaalit toistettiin perinteisellä DirAC-tekniikalla menetelmässä *normDirAC*. Koehenkilöiden tehtävänä oli arvostella toiston yleistä laatua verrattuna referenssiin. Kysymyksenä oli: ”Arvioi havaittu laadun heikkeneminen”. Arvosana 5 tarkoitti, ettei havaitse eroa, ja 1 tarkoitti, että ero on erittäin häiritsevää. Kuuntelukokeiden tulokset ovat nähtävissä kuvassa 4. Voidaan nähdä, että *spacedDirAC*-menetelmän laatu on verrattain hyvä kaikilla näytteillä. Tulee huomata, että testissä verrattiin viisikanavaisia menetelmiä kolmetoistakanavaisen referenssiin. Lisäksi voidaan nähdä, että esitetty menetelmä on havaitulta laadultaan parempi tai yhtä hyvä kaikilla näytteillä verrattuna sekä perinteiseen DirAC-toistoon että myös erillismikrofonisignaalien suoraan toistoon.

## 6 DISKUSSIO

Esitettyä menetelmää voidaan ajatella perinteisten erillismikrofonitekniikoiden parametrisena parannuksena. Se osa äänestä mikä mikrofonisignaaleissa toimii suoraan toistettuna sellaisenaan, toistetaan sellaisenaan. Vain ongelmakohtia pyritään käsittelemään. Matalilla taajuuksilla mikrofonisignaalit ovat yleensä liian koherentteja, joten ne voidaan dekorreloida niiltä osin. Lisäksi lokalisoitavissa olevat osat äänestä voidaan toistaa amplitudipanoroinnilla, mikä lisää suuntahavainnon tarkkuutta ja pistemäisyyttä.

Esitettyä menetelmää voidaan ajatella myös jälkikäsitteilytyökaluna. Äänityksen voi tehdä haluamallaan tekniikalla, ja jos lisäksi haluaa parantaa lokalisaatiota, voi käyttää esitettyä menetelmää. Prosessoinnin voi ottaa käyttöön saumattomasti, koska menetelmällä prosessoitua ja prosessoimatonta ääntä voi miksata vapaasti valittavassa suhteessa. Täten miksaaja voi vapaasti valita haluamansa prosessoinnin määrän kuuntelemalla.

## 7 YHTEENVETO

Artikkelissa esitettiin DirAC-tekniikkaan pohjautuva menetelmä erillismikrofoniäänitusten parametriseen käsittelyyn. Menetelmässä analysoidaan taajuuskaistoittain äänilähteiden suunta sekä diffuusisuus, minkä avulla toisto pyritään tekemään optimaalisesti. Menetelmä parantaa havaittua laatua perinteisiin erillismikrofonitekniikoihin verrattuna tarjoamalla tarkempaa ja stabiilimpaa äänilähteiden lokalisaatiota, kuitenkin siten, että hyvä tilantuntu säilyy. Erillismikrofoniäänitusten alhaisen mikrofoniien välisen koherenssin takia voidaan parantaa myös perinteisen DirAC-tekniikan laatua.

## 8 KIITOKSET

Kirjoittaja haluaa kiittää Archontis Politista avusta artikkelin tekemisessä. Fraunhofer IIS ja GETA-tutkijakoulu ovat tukeneet tätä tutkimusta.

## VIITTEET

- [1] Francis Rumsey. *Spatial Audio*. Focal Press, Oxford, 2001.
- [2] Michael A. Gerzon. Ambisonic in multichannel broadcasting and video. *J. Audio Eng. Soc.*, 33(11):859–871, November 1985.
- [3] Ville Pulkki. Spatial sound reproduction with directional audio coding. *J. Audio Eng. Soc.*, 55(6):503–516, June 2007.
- [4] Ville Pulkki. Microphone techniques and directional quality of sound reproduction. In *AES 112th Convention*, Munich, Germany, 2002.
- [5] Mikko-Ville Laitinen, Fabian Kuech, and Ville Pulkki. Using spaced microphones with directional audio coding. In *AES 130th Convention*, London, UK, May 2011.
- [6] Archontis Politis, Mikko-Ville Laitinen, Jukka Ahonen, and Ville Pulkki. Parametric spatial audio coding for spaced microphone array recordings. In *AES 134th Convention*, Rome, Italy, May 2013.
- [7] Archontis Politis, Mikko-Ville Laitinen, and Ville Pulkki. Spatial audio coding with spaced microphone arrays for music recording and reproduction. In *21st International Congress on Acoustics*, Montreal, Canada, June 2013.
- [8] Günther Theile. Natural 5.1 music recording based on psychoacoustic principles. In *AES 19th International Conference*, Schloss Elmau, Germany, June 2001.
- [9] W. Gary Elko. Spatial Coherence Functions for Differential Microphones in Isotropic Noise Fields. In Michael Brandstein and Darren B. Ward, editors, *Microphone Arrays*, chapter 3, pages 61–85. Springer-Verlag, Berlin, Heidelberg, New York, 2004.