

TELEKONFERENSSISOVELLUS JA SUUNTAMIKROFONITEKNIikka DIRAC-MENETELMÄLLE

Jukka Ahonen, Ville Pulkki

Akustiikan ja äänenkäsittelytekniikan laboratorio (TKK)
PL 3000, FI-02015 TKK
jukka.ahonen@acoustics.hut.fi

1 JOHDANTO

DirAC (Directional Audio Coding) on tiläänen äänittämiseen sekä toistamiseen kehitetty menetelmä. Se mahdollistaa korkeatasoisen äänentoiston mielivaltaisella äänentoistojärjestelmällä, suunta- ja tilainformaation välittämisen telekommunikaatiojärjestelmässä monokanavan sivukaistalla alhaisella siirtonopeudella, sekä stereon konvertoinnin monikanavaiseksi. DirAC on samalla myös suuntamikrofoniteknikka, jota voidaan käyttää mielivaltaisten suuntakuvioiden luomiseen.

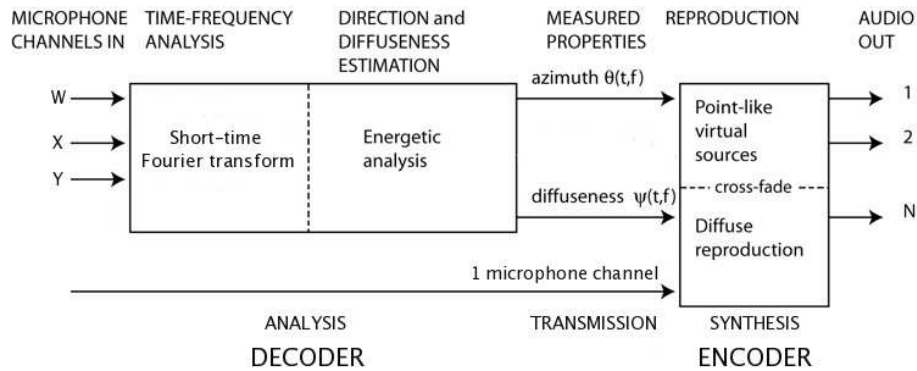
DirAC perustuu samoihin periaatteisiin ja osittain samoihin menetelmiin kuin huoneakustiikan toistoon tarkoitettu SIRR-menetelmä (Spatial Impulse Response Rendering) [1] [2]. Äänikenttää ei pyritä toistamaan täydellisesti, vaan DirAC-prosessoinnissa otetaan huomioon vain ihmisen kuulon kannalta relevantit äänikentän ominaisuudet. Korvien välisten aika- ja tasoerojen (ITD ja ILD), korvien välisen koherenssin (IC) ja äänen värin katsotaan toistuvan oikein, mikäli äänen tulosuunta, diffuusisuus ja spektri toistetaan oikein. DirAC-prosessointi suoritetaan ihmiskorvan aika- ja taajuusresoluution tarkkuudella.

Tässä paperissa käsitellään yksinkertaista DirAC-versiota. Korkealaatuisen tiläänen prosessointiin tarkoitettu versio on esitelty viitteessä [3].

2 YKSINKERTAISEN DIRAC-VERSION PERIAATE

Yksinkertaista DirAC-versiota käytetään telekonferenssi- ja suuntamikrofonisovelluksissa. Kuvassa 1 on esitetty lohko-kaavio telekonferenssisovellukselle. Alkuun mikrofonisignaaleille suoritetaan aika-taajuusanalyysi, ja äänen tulosuunta sekä diffuusisuus estimoidaan ajan funktiona käyttäen energia-analyysia DirAC-dekooderissa. Tämän jälkeen suunta- ja tilainformaatio siirretään monokanavan sivukaistalla DirAC-enkooderille. Synteesivaiheessa ääni jaetaan ei-diffuusiin sekä diffuusiin ääneen, jotka toistetaan eri tekniikoin halutulla määrällä kaiuttimia tai kuulokkeilla. Tyypillisessä telekonferens-

sineuvottelussa kaiuttimet on aseteltu ainoastaan vaakatasoon, jolloin korkeuskulmaa äänen tulosuunnalle ei tarvitse estimoida eikä siirtää. Telekonferenssisovelluksessa prosessointi suoritetaan puhelinkaistalla 300-3400 Hz:n välillä.



Kuva 1: Lohkokaavio telekonferenssisovellukselle

Äänen tulosuunnan ja diffuusisuuden analysoinnissa tarvittavat äänikentän suuret saadaan mitattua B-formaattimikrofonilla. Se koostuu pallomikrofonista (nollannen asteen painemikrofoni) sekä dipolimikrofoneista (ensimmäisen asteen painegradienttimikrofoni). B-formaattiaänityksessä äänikenttä jaetaan palloharmonisiin, jolloin pallomikrofonisignaali W vastaa nollannen asteen harmonista ja dipolimikrofonisignaalit X , Y ja Z vastaavat ensimmäisen asteen harmonisia x -, y - ja z -koordinaattiakseleille. Korkeampi aste palloharmonisessa lisää sen suuntavuutta.

DirAC ratkaisee useita monikanavaisessa äänentoistossa sekä telekommunikaatiossa esiintyviä ongelmia. Perinteisesti monikanavaisissa äänityksissä käytetään yhtä montaa suuntamikrofonia kuin toistossa on kaiuttimia, jolloin käytettävä äänentoistojärjestelmä tulee olla jo etukäteen tiedossa. Monikanavaisessa telekommunikaatiossa useampi mikrofonisignaali vaatii enemmän siirtokaistaa kuin pelkän monoäänien siirto. DirAC-telekonferenssisovelluksessa suunta- ja tilainformaation siirto monikanavan sivukaistalla lisää vain vähän siirtokaistaa.

2.1 Energia-analyysi

Energia-analyysi perustuu ääni-intensiteettiin sekä äänikentän energiatiheuteen. Intensiiteetti kuvaa energian virtaamista äänikentässä ja energiatiheys energian määrää äänikentän pisteessä. Näiden energia-analyysissä tarvittavien suureiden laskentaan vaadittavia kenttäsuureita ovat äänipaine ja hiukkasnopeus. B-formaattimikrofonin pallomikrofonisignaali edustaa äänipainetta ja dipolimikrofonisignaalit hiukkasnopeutta.

Yksinkertaisessa DirAC-versiossa energia-analyysi lasketaan kullakin ajanhetkellä kul-

lekin taajuuskomponentille käyttäen lyhytaikaista Fourier-muunnosta (STFT). Mikrofonisignaalien STFT-spektreistä lasketaan äänen hetkellinen intensiteetti x - ja y -akseleille äänipaineen ja hiukkasnopeuden tulona kaavasta

$$\begin{aligned} I_X(t, f) &= \frac{1}{\sqrt{2}Z_0} \operatorname{Re}\{W^*(t, f) \cdot X(t, f)\} \\ I_Y(t, f) &= \frac{1}{\sqrt{2}Z_0} \operatorname{Re}\{W^*(t, f) \cdot Y(t, f)\}, \end{aligned} \quad (1)$$

missä Z_0 vastaa ilman akustista impedanssia, t on aika ja f on taajuus. Akustinen impedanssi lasketaan $Z_0 = \rho_0 \cdot c$, missä ρ_0 on ilman keskitiheys ja c on äänennopeus. Hetkellinen ääni-intensiteetti $I(t, f)$ vastaa todellista eli aktiivista intensiteettiä, jolloin sen laskennassa huomioidaan ainoastaan STFT-spektrien reaali-osat. Imaginaariosa vastaa reaktiivista intensiteettiä, joka on nolla tasoallolle.

Äänen hetkellinen tulosuunta on päinvastainen energiavirtauksen suunnalle, joten se voidaan laskea hetkellisestä intensiteetistä $I(t, f)$. Atsimuuttikulma $D(t, f)$ asteina lasketaan kaavasta

$$D(t, \omega) = \begin{cases} \arctan \left[\frac{-I_Y(t, f)}{-I_X(t, f)} \right] & , \text{ jos } I_Y(t, f) \geq 0 \\ \arctan \left[\frac{-I_Y(t, f)}{-I_X(t, f)} \right] - 180^\circ & , \text{ jos } I_Y(t, f) < 0. \end{cases}$$

Äänen hetkellinen energiatiheys $E(t, f)$ voidaan laskea kaavasta

$$E(t, f) = \frac{1}{2} \rho_0 Z_0^{-2} \left[|W(t, f)|^2 + \frac{|X(t, f)|^2 + |Y(t, f)|^2}{2} \right]. \quad (2)$$

Äänen diffuusisuus $\psi(t, \omega)$ saadaan hetkellisen intensiteetin ja energiatiheyden suhteesta kaavalla

$$\psi(t, f) = 1 - \frac{\sqrt{2} |\operatorname{Re}\{W^*(t, f) \cdot [X(t, f) + Y(t, f)]\}|}{|W(t, f)|^2 + \frac{|X(t, f)|^2 + |Y(t, f)|^2}{2}}. \quad (3)$$

2.2 Synteesi

DirAC-synteesissä ääni jaetaan ei-diffuusiin ja diffuusiin osaan, jotka toistetaan soveltuvin tekniikoin. Ei-diffuusit virtuaaliset äänilähteet tuotetaan telekonferenssisovelluksessa käyttäen kaksiulotteista amplitudipanorointia sekä tangenttilakia, jolla lasketaan estimoituja atsimuuttikulmia vastaavat panorointikertoimet.

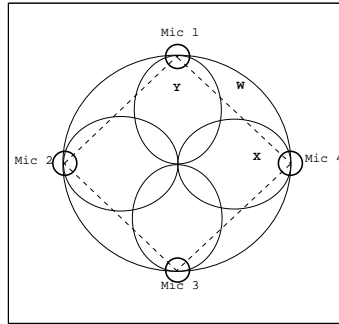
Äänen analysoitu tulosuunta saattaa vaihdella nopeasti ajassa, jolloin toistettavassa äänessä esiintyy kuultavia artefakteja panorointikertoimien muuttuessa nopeasti. Näistä

päästään eroon ajallisella keskiarvoituksella, jossa kaiuttimien panorointikertoimia hidastetaan. Hidastus toteutetaan ensimmäisen asteen IIR-suotimella.

Telekonferenssisovelluksessa diffuusi ääni syntetisoidaan käyttäen satunnaista amplitudipanorointia, jossa diffuusin äänen suunta vaihtelee satunnaisesti eri ajanhetkillä ja taajuuksilla ollen epäkoherenttia kaiuttimien välillä. Myös diffuusille äänelle tulee suorittaa panorointikertoimien hidastus.

3 B-FORMAATIN MUKAINEN MIKROFONIHILA

B-formaatin mukaiset signaalit voidaan tuottaa koinsidenssitekniikalla, jossa tietty määrä mikrofoneja asetellaan lähemmäksi toisiaan [4]. Kaupallisissa sovelluksissa signaalien luomiseen käytetään kohtuullisen kalliita ensimmäisen asteen mikrofoneja [5]. B-formaatin mukaiset signaalit voidaan kuitenkin luoda käyttämällä myös halvempia nolannen asteen mikrofoneja [6].



Kuva 2: B-formaatin mukainen mikrofonihila.

Telekonferenssi- ja suuntamikrofonisovelluksessa B-formaatin mukaiset signaalit muodostetaan mikrofonihilalla, joka koostuu neljästä pallomikrofonista (kuva 2). Dipolimikrofonisignaalit X ja Y luodaan vähentämällä hilan vastakkaisiin kulmiin sijoitettujen mikrofonien signaalit toisistaan kaavalla

$$\begin{aligned}
 X(t, f) &= -\frac{j \cdot c}{f_c \cdot d} [W_1(t, f) - W_3(t, f)] \\
 Y(t, f) &= -\frac{j \cdot c}{f_c \cdot d} [W_2(t, f) - W_4(t, f)],
 \end{aligned} \tag{4}$$

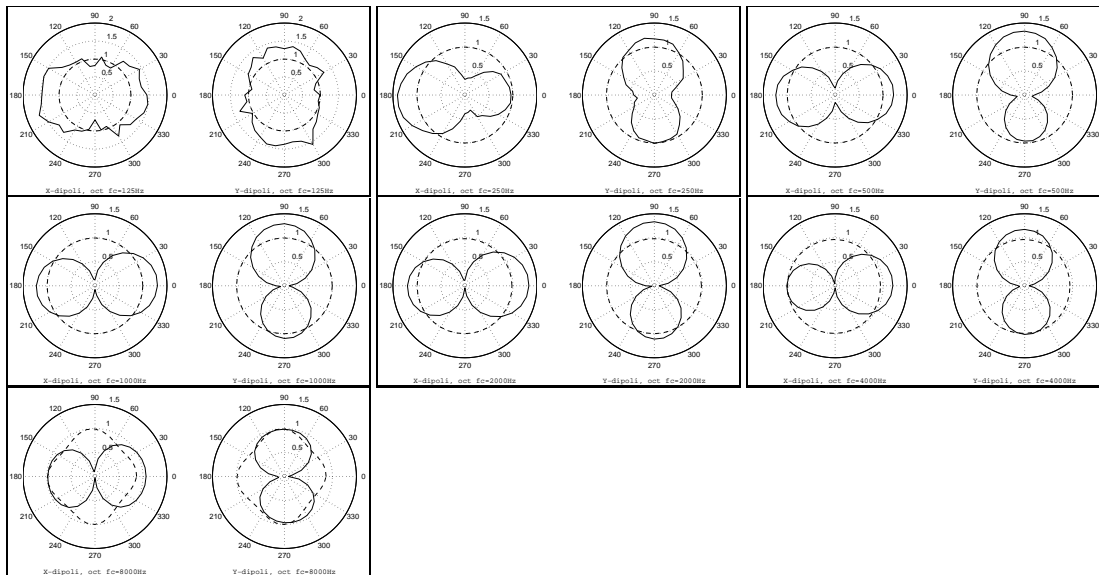
missä $W_{1..4}(t, f)$ on mikrofonisignaalin STFT-spektri, j on imaginaariyksikkö, f_c on keskitaajuus taajuuskomponentille ja d on mikrofonien välinen etäisyys. Pallomikrofo-

nisignaali W saadaan laskemalla hilan mikrofoniin signaalit yhteen ja jakamalla summa neljällä. Painemikrofoni ja dipolimikrofonit mittaavat samaa äänikentän pistettä.

Dipoliin X ja Y suuntakuviot ovat taajuusriippuvaisia. Ylärajataajuus dipoleille määräytyy pallomikrofonien välisen etäisyyden perusteella. Ääniaallon puolikkaan ollessa lyhyempi kuin mikrofoniin välinen etäisyys, esiintyy spatiaalista laskostumista. Alarajataajuus määräytyy pallomikrofonien sisäisen kohinan perusteella. Kun mikrofoniin signaalit vähennetään toisistaan, niin pienillä taajuuksilla signaalit kumoavat lähes kokonaan toisensa ja yksittäisten mikrofoniin kohinat summautuvat lisäten dipoliin sisäistä kohinaa 3 dB.

Dipoliin verrattain suuri kohina rajoittaa tyypillisesti mikrofoniin hilan käyttöä. DirAC:ssa dipoleita käytetään kuitenkin ainoastaan suunnan ja diffuusisuuden analysointiin, ei niiden tuottamia audiosignaaleita.

Kuvassa 3 on esitetty B-formaatin mukaisen mikrofoniin hilan suuntakuviot oktaavikaistoittain. Mittaus suoritettiin kymmenen asteen välein kaiuttomassa huoneessa. Testisignaalinä käytettiin logaritmista taajuuspyyhkäisyä ($L_{Aeq} = 65$ dB) ja äänilähteenä Genelec 8030A kaiutinta (etäisyys 1,5 m). Vastakkaisten pallomikrofonien (Sound Professional BMC-12) välinen etäisyys oli 2 cm, jolloin laskennallinen ylärajataajuus dipoleille on noin 8,5 kHz. Pallomikrofonien taajuusvasteet on ekvalisoitu. Viitteissä [7] ja [8] on esitetty vastaavanlaiset suuntakuviomittaukset kaupallisille B-formaattimikrofoneille.



Kuva 3: Suuntakuviot B-formaatin mukaiselle mikrofoniin hilalle oktaavikaistoittain mitattuna. Dipolit (yhtenäinen viiva) on normalisoitu pallomikrofoniin signaalin W (katkoinen viiva) suhteen.

Suuntakuvioista nähdään, että keskitaajuuksilla 500-8000 Hz X - ja Y -dipolit ovat suuntaavia ja tarkkoja, joskin hieman epäsymmetrisiä. Keskitaajuudella 8000 Hz pallosignaali W muuttuu suuntaavammaksi spatiaalisesta laskostumisesta johtuen. Dipolit ovat kohtuullisen tarkkoja ja suuntaavia vielä 250 Hz:n keskitaajuudella. Keskitaajuudella 125 Hz dipolit muuttuvat enemmän ympäriseiteleviksi, mutta silti niissä on havaittavissa kiilamainen muoto.

B-formaatin mukainen mikrofonihiila ja DirAC-prosessointi muodostavat suuntamikrofonin, jolla voidaan luoda mielivaltaisia suuntakuvioita. Halutusta suunnasta saapuvaa ääntä vahvistetaan ja toisaalta ääni, joka saapuu ei-halutusta suunnasta, voidaan vaimentaa.

4 KIITOKSET

Tämä työ on Suomen Akatemian rahoittama (projekti no. 105780).

VIITTEET

- [1] MERIMAA J & PULKKI V, Spatial Impulse Response Rendering I: Analysis and synthesis, *J. Audio Eng.Soc.*, **53**(2005), 1115–1127.
- [2] PULKKI V & MERIMAA J, Spatial Impulse Response Rendering II: Reproduction of diffuse sound and listening tests, *J. Audio Eng.Soc.*, **54**(2006), 3–20.
- [3] PULKKI V, Directional Audio Coding in spatial sound reproduction and stereo up-mixing, in *AES 28th Conference, Piteå, Sweden*, (2006).
- [4] JULSTROM S, An intuitive view of coincident stereo microphones, *J. Audio Eng.Soc.*, **39**(1991), 632–649.
- [5] BENJAMIN E & CHEN T, The native B-format microphone: Part I, in *AES 119th Convention, New York, USA, preprint 6621*, (2005).
- [6] MERIMAA J, Applications of a 3-D microphone array, in *AES 112th Convention, Munich, Germany, preprint 5501*, (2002).
- [7] AHONEN J, PULKKI V, & LOKKI T, Teleconference application and B-format microphone array for Directional Audio Coding, in *AES 30th Conference, Saariselkä, Finland*, (2007).
- [8] FARINA A, Anechoic measurement of the polar plots of a Soundfield MKV B-format microphone, *Technical report, Universita Degli Studi Di Parma*.