

Äänilähteen reaaliaikainen paikannus mikrofoniryhmän avulla

TURUN YLIOPISTO
Tietotekniikan laitos
LuK-tutkielma
Tietojenkäsittelytiede
Kesäkuu 2024
Tuomas Lundberg

TURUN YLIOPISTO
Tietotekniikan laitos

TUOMAS LUNDBERG: Äänilähteen reaaliaikainen paikannus mikrofoniiryhmän avulla

LuK-tutkielma, 23 s.
Tietojenkäsittelytiede
Kesäkuu 2024

Digitaalinen signaalinprosessointi on tullut viime vuosikymmeninä hyvin paljon tehokkaammaksi ja halvemmaksi, ja näin ollen myös monet äänilähteen paikantamiseen liittyvän insinööriyön haasteet ovat helpottuneet huomattavasti. Tämä muutos on mahdollistanut lukuisten äänilähteen paikantamissovellusten kaupallisen toteuttamiskelpoisuuden. Tähän joukkoon teknologioita liittyy kuitenkin lukuisia algoritmisiä sekä optimointihaasteita, jotka on otettava huomioon, jotta sovellusta suunniteltaessa osattaisiin valita sopiva paikannusmenetelmä.

Tutkielma perehtyy viimeisen noin kymmenen vuoden aikana alalla tapahtuneisiin edistysaskeleisiin kirjallisuuskatsauksen muodossa. Ensin luodaan yleisluontoinen katsaus äänilähteen paikantamiseen fysikaalisena, matemaattisena sekä laskennallisena ongelmana. Sitten perehdytään alan tyyppisimpiin tekniikoihin sekä näiden etuihin ja ongelmiin. Lopuksi vertaillaan alan hiljattain julkaistusta kirjallisuudesta löytyviä lähestymistapoja ja tekniikoita, näihin liittyviä haasteita sekä niiden tuloksia.

Äänilähteen paikantamisen viimeaikaisessa kehityksessä on havaittavissa IoT:n kasvun myötä selkeä trendi autonomisten ja mobiilien järjestelmien konekuulon saralla. Myös monet vakiintuneet teknologiat ovat edistyneet ja sofistikoituneet, ja koneoppimismetodien yhdistäminen perinteisiin signaalinkäsittelymenetelmiin on yleistynyt. Laskentatehon halpeneminen lienee myös kasvattanut laskentavaatimuksiltaan raskaampien menetelmien suosiota.

Asiasanat: ääni, akustiikka, paikannus, mikrofoni, mikrofoniverkko, mikrofoniiryhmä, SSL, DOA

Sisällys

1	Johdanto	1
2	Äänen paikantamisen perusteet	4
2.1	Akustinen ääni	4
2.2	Äänen taltionti ja digitaalinen ääni	5
2.3	Äänen paikantaminen	6
3	Yleiskatsaus paikannusmenetelmiin	10
3.1	Paikannuksen lähtökohdat ja ongelmakuvaus	10
3.2	Aika-akseliin perustuvat tekniikat	13
3.3	Spektrianalyysiin perustuvat tekniikat	15
4	Menetelmien vertailua	17
4.1	Laskentakustannukset ja IoT	17
4.2	Sovellusympäristön tuomat haasteet	19
4.3	Koneoppimisen vaikutus	20
5	Yhteenveto	22
	Lähdeluettelo	24

Termistö

AVS Acoustic Vector Sensor

CNN Convolutional Neural Network

DOA Direction Of Arrival

GCC-PHAT Generalised Cross-correlation Phase Transform

GCC Generalised Cross-correlation

IoT Internet of Things

MLP Multi-layer Perceptron

MUSIC Multiple Signal Classification

SNR Signal-to-noise Ratio

SRP-PHAT Steered Response Power Phase Transform

SRP Steered Response Power

SSL Sound Source Localisation

TDOA Time Difference Of Arrival

1 Johdanto

On selvää, että ääntä voi paikantaa. Äänilähteen paikantamiseen tarvittavan informaation on *pakko* olla olemassa, koska eläimet kykenevät paikantamaan äänilähteitä. Lisäksi kahden korvan kehittyminen osoittaa, että ääntä aistivia elimiä tarvitaan äänen paikannukseen luultavasti enemmän kuin yksi. Kuten monen muunkin teknologian tapauksessa, tässäkin luonnon tarjoama biologinen esimerkki on paitsi validoinut ongelman olevan ratkaistavissa, myös tarjonnut ihmiselle inspiraatiota siihen, miten sitä voisi lähestyä.¹

Äänen tehokkaalle paikantamiselle on myös lukemattomia hyödyllisiä sovelluksia. Autonomiset robotit ja kulkuneuvot voivat hyödyntää äänilähteen paikantamista vaikka navigaation tai ihmisen kanssa luontevasti kommunikoinnin apuna. Paikannusta voidaan hyödyntää myös turvallisuus- ja valvonta-alalla sekä ase- ja puolustusalalla, esimerkiksi aseiden laukauksen tai ikkunan hajoamisen suunnan tunnistuksessa. Teknologiaa voi hyödyntää myös kuulolaitteissa sekä äänen vahvistamisessa paikoissa, joissa perinteisten mikrofoniin käyttäminen ei ole mahdollista tai toivottavaa: esimerkiksi luentosalissa mielivaltaisessa paikassa istuvan puhujan äänen vahvistus on mahdollista toteuttaa äänenpaikannusteknologialla. Vastaavia mahdollisia sovelluksia on myös telekonferenssialalla.

Tutkimusyhteisön oli luontevaa lähestyä äänen paikantamista tietojenkäsittelyn ja elektroniikan avulla, kun nämä alat olivat kehittyneet niin pitkälle, että äänen

¹Asiaa voisi verrata vaikkapa ilmailun kehittymiseen.

paikantaminen alkoi vaikuttaa realistiselta toteuttaa tai prototypoida matemaattisen mallintamisen ja algoritmiikan avulla. Äänen paikantamista on tosin tehty jo ennen kuin nämä alat olivat kovin pitkälle kehittyneitä, esimerkiksi kaikuluotauksen ja akustisten peilien muodossa [1].

Tässä tutkielmassa tutustutaan *akustisen* eli fysikaalisessa väliaineessa liikkuvan ääniaallon paikantamiseen. Luvussa 2 tutustutaan akustisen äänilähteen paikantamisen perusteisiin, minkä jälkeen luvuissa 3 ja 4 selvitetään alan nykytilaa, viimeaikaisia edistysaskeleita sekä näihin liittyviä haasteita ja rajoitteita. Rajaus tehdään tiukasti mikrofoni ryhmää tai -verkkoa (engl. *microphone array* tai *microphone grid*) hyödyntäviin, passiiviseen paikannukseen perustuviin teknologioihin. Tutkielmasta rajataan siis pois akustisiin vektoriantureihin sekä kaikuluotaukseen perustuvien teknologioiden käsittely. Akustiset vektorianturit (AVS, engl. *Acoustic Vector Sensor*) hyödyntävät paikantamiseen ääniaallossa tapahtuvan hiukkasliikkeen suuntaa ja kykenevät näin suoraan osoittamaan ääniaallon suunnan. Kaikuluotaukseen perustuvissa teknologioissa taas lähetetään ääniaalto, jonka heijastumista voidaan laskea kappaleiden suuntimat ja muodot.

Monet äänen paikantamisen sovelluskohteet vaativat likimain reaaliaikaisuutta, mikä asettaa huomattavia rajoitteita teknologioiden prosessointiaikavaatimuksille. Mobiileissa sovelluksissa, kuten esimerkiksi autonomisissa roboteissa, laskentaresurssit ovat usein myös harvassa. Näiden rajoitteiden takia painopiste on tekniikoissa, jotka ovat laskentavaatimuksiltaan sovellettavissa käyttötapauksiin, joissa paikannuksen tulee olla käytännössä reaaliaikaista.

Tutkielma on toteutettu kirjallisuuskatsauksena. Kirjallisuuskatsauksen aineisto on täysin peräisin IEEE Xplore[®] -tietokannasta. Haussa käytettiin seuraavaa hakulauseketta: (*“real-time” OR “real time”*) AND (*“sound source” OR “acoustic source”*) AND (*localization OR localisation*) AND *microphone** AND (*grid OR array*).

Haku tehtiin perjantaina 8.3.2024. Aineistoon otettiin mukaan vain aikakausijulkaisuissa julkaistuja artikkeleita, jotka on julkaistu aikaisintaan 2013. Tällä hakurajauksella hakutuloksia oli 23. 13 artikkelia jätettiin abstraktin perusteella ja/tai muun harkinnan varaisesti käsittelyn ulkopuolelle. Lisäksi aineistona on käytetty alan peruskirjallisuutta (ks. lähteet [2], [3]) sekä muita alan kehitykselle tärkeitä julkaisuja (ks. lähteet [4]–[8]).

2 Äänen paikantamisen perusteet

Jotta voitaisiin kokonaisvaltaisesti ymmärtää äänilähteen paikantamisen luonnetta insinööriongelmana, on tarkasteltava ongelman viitekehystä muutamasta eri näkökulmasta: tulee ymmärtää ääni sekä fysikaalisena ilmiönä että abstraktina signaalina, sekä lisäksi näiden kahden rajapinnassa tapahtuvan äänen taltioimisen tai toisintamisen reunaehdot.

2.1 Akustinen ääni

Akustinen ääni on väliaineessa etenevää, kuultavissa olevaa aaltoliikettä [9]. Ääniaalto on väliaineen tasapainotilan dynaaminen häiriö, joka etenee, kun tasapainotilasta poikkeava kohta väliaineessa pyrkii korjaamaan itsensä takaisin tasapainotilaan [2, p. 1]. Rummun kalvon värähtäessä ulospäin rummusta juuri kalvon pinnalla ollut ilma joutuu väistymään rumpukalvon tieltä, mikä aiheuttaa paikallisesti korkeamman ilmanpaineen tämän ilman pakkautuessa ulospäin rumpukalvosta. Tämä ylipaine pyrkii purkautumaan niihin suuntiin, joihin se on vähiten työläintä, eli tässä tapauksessa kaikkialle paitsi takaisin rumpukalvoa kohti. Näin ollen ilmanpaine tasaantuu pois päin rummusta. Nyt hieman kauempana rumpukalvosta on kuitenkin puolestaan taas pieni ylipaine, joka jälleen pyrkii purkautumaan vähiten työläiseen suuntaan. Mitä kauempana ylipaine on rumpukalvosta, sitä laajemmalle alueelle paineaalto on edennyt, ja täten myös sitä laajemmalle alueelle paineaallon energia

on jakautunut. Ketjureaktio siis etenee ja heikkenee, kunnes se hukkuu ympäristön muiden äänien joukkoon.

Edellinen yksinkertaistettu esimerkki jättää käsittelemättä lukemattomia äänen fysikaalisia ominaisuuksia, kuten ääniaaltojen heijastumista ja diffraktiota, väliaineen ominaisuuksia sekä jaksottaista aaltoliikettä, joka äänen muodostumiseen usein liittyy. Olennaista on kuitenkin se, että akustinen ääni on aina aaltona etenevä painero.

Alussa esitetty määritelmä tarkentaa äänen olevan kuultavissa. Kuulemisen käsite on kuitenkin määriteltävä, jotta esitettyä määritelmää voidaan käyttää. Keskimääräinen ihmiskorva kuulee ääniaaltoja, joiden taajuus on välillä 20 – 20 000 Hz, eli joissa ilmanpaine kasvaa ja pienenee 20 – 20 000 kertaa sekunnissa [2, p. 316]. Kuitenkin monen eläimen kuuloalue on tätä huomattavasti laajempi, ja myös mikrofonin voisi sanoa “kuulevan” kaikki sellaiset ääniaallot, jotka kykenevät tuottamaan mikrofonissa mitattavissa olevan, sähköisen signaalin.

2.2 Äänen taltionti ja digitaalinen ääni

Mikrofonin kalvon voidaan ajatella olevan äänilähteen aiheuttajan vastakohta. Kun ääniaallon positiivisen ilmanpaineen alue saavuttaa mikrofonin kalvon, ilmanpaine työntää kalvoa pois päin itsestään. Vastaavasti negatiivinen ilmanpaineen alue vetää kalvoa itseään kohti. Tämä kalvon liike indusoi mikrofonissa sähkövirran, jonka suuruus riippuu suoraan kalvon siirtymästä. Ääniaalto siis muuntuu mikrofonissa sähköiseksi signaaliksi, jolla on kutakuinkin samat ominaisuudet kuin ääniaallolla. [2, p. 286] Kun ääntä tallennetaan digitaalisesti, sähköinen signaali kulkee mikrofonista esivahvistuksen kautta niin sanottuun *A/D-muuntimeen* (engl. *A/D converter*, sanoista *Analog* ja *Digital*), jossa analoginen, jatkuva signaalifunktio voidaan muuntaa tietyllä aika- ja amplitudiresoluutiolla tallennettavaksi digitaaliseksi signaaliksi. Esimerkiksi CD-levyyn tallennettu äänisignaali on 16-bittinen, ja sen

näytteenottotaajuus on 44 100 Hz. Tällaisella resoluutiolla äänitetty audiosignaali on taltioitu 44 100 kertaa sekunnissa siten, että jokaisella näistä 44 100 kerrasta analogisen, sähköisen signaalin amplitudi on muunnettu 16 bitin tarkkuudella numeroksi. [10] 16 bitillä voidaan esittää $2^{16} = 65\,536$ erilaista lukua ja siten myös amplitudia, ja suurin osa CD-levyjä kuulleista ihmisistä kokee tämän resoluution olevan riittävän suuri autenttiseen äänen takaisintoistoon.

44 100 hertsin näytteenottotaajuusstandardi ei ole täysin mielivaltaisesti päätetty. Harry Nyquist osoitti vuonna 1928, että alkuperäistä vastaavan analogisen signaalin uudelleenmuodostamiseksi digitaalisesta signaalista tulee näytteenottotaajuuden olla yli kaksinkertainen alkuperäisen signaalin korkeimpaan osataajuuteen nähden [4]. Näin ollen 44 100 hertsin näytteenottotaajuudella kyetään tallentamaan ja toisintamaan akustista ääntä $\frac{44\,100}{2} = 22\,050$ hertsiin asti, mikä riittää kattamaan ihmisen koko kuuloalueen.

Jotta digitaalisesti tallennettu äänisignaali voidaan muuntaa takaisin akustiseksi ääneksi, signaalin on läpikäytävä äänen taltioimiseen nähden käänteinen prosessi. Tällöin digitaalinen signaali muunnetaan *D/A-muuntimen* (*Digital to Analog*) avulla analogiseksi, sähköiseksi signaaliksi, ja tällä signaalilla voidaan liikuttaa vaikka kaiuttimen elementtiä ja näin muuntaa sähköinen signaali jälleen ilmanpaineen vaihteluiksi eli akustiseksi ääneksi [2, p. 398].

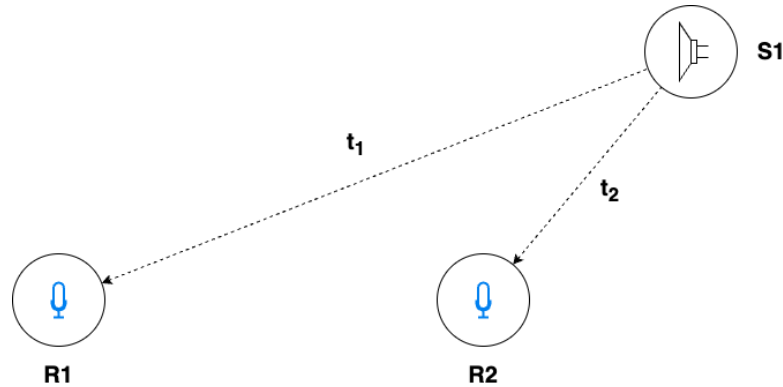
2.3 Äänen paikantaminen

Suurin osa eläinlajeista on kehittynyt aistimaan ääntä jossain muodossa, ja monet erittäin tarkasti. Toisin kuin elektromagneettinen säteily eli valo, jonka avulla näköaisti toimii, ääniaallot kykenevät taivuttamaan kulmien taakse sekä monien kiinteiden esteiden läpi makroskooppisessa mittakaavassa [2, p. 417]. Äänen siis kykenee aistimaan näköesteiden läpi. Tämä on hyvin ilmeisistä syistä osoittautunut evoluutiivisesti erittäin merkittäväksi kilpailuvaltiksi. Pitkässä nurmikossa väijyvä peto tai

kaukaa lähestyvä lauma saaliseläimiä saattaisivat jäädä huomaamatta ilman tarkkaa kuuloaistia. Eläinten kuuloaisti toimii hyvin pitkälti samalla periaatteella kuin mikrofoni: ääniaallot liikuttavat korvassa sijaitsevaa tärykalvoa, ja tästä aiheutuva sisäkorvan painenvaihtelu muuntuu kuuloreseptoreissa hermoimpulssiksi [11].

Systemiä, joka kykenee muuntamaan ääniaallon hyödynnettäväksi informaatioksi, kuten esimerkiksi mikrofonia tai korvaa, kutsutaan *muuntajaksi* (engl. *transducer*) [2, p. 390]. Käytännössä tässä tutkielmassa puhutaan kuitenkin *vastaanottimesta* (engl. *receiver*), sillä kontekstina on lähes poikkeuksetta ääniaallon saapuminen muuntajaan. Yksittäisen vastaanottimen havainnoimaa ääntä kutsutaan *yksikanavaiseksi ääneksi* [12]. Yhden äänikanavan informaatio on kuitenkin 1-ulotteista; jokaisella ajanhetkellä signaalin määrittää yksikäsitteisesti sen amplitudi. Tästä johtuen pelkän äänenpaineen vaihtelun avulla ei yksikanavaisesta äänestä ole lähtökohteisesti mahdollista havainnoida äänen tulokulmaa; mistä tahansa suunnasta saapuva ääni voi teoriassa aiheuttaa samanlaisen äänihavainnon.

Jotta äänen tulokulmaa voidaan havainnoida, tarvitaan useampi vastaanotin. Tästä syystä ääntä aistiville eläimille on kehittynyt kaksi erillistä kuuloelintä. Samasta syystä äänilähteet usein taltioidaan ja toisinnetaan vähintään *stereona* eli kahta kanavaa käyttäen, jotta ääntä toistettaessa kyetään luomaan illuusio äänen tulokulmasta [12]. Kaksi vastaanotinta mahdollistaa äänen tulokulman määrittämisen, koska ääniaallon etenemisnopeus väliaineessa on rajallinen (ks. kuva 2.1). Ääniaallot etenevät merenpinnan tason ilmanpaineessa ja 20 celsiusasteen lämpötilassa noin 343 m/s nopeudella [13], siis noin 34 senttimetriä millisekunnissa. Koska korvien välinen etäisyys on suurimmalla osalla ihmisistä noin 14–17 senttimetriä, täysin vasemmalta saapuva ääni saavuttaa ihmisen vasemman korvan noin puoli millisekuntia ennen oikeaa. Koska ääniaalto on molempiin korviin saapuessaan aaltomuodoltaan suunnilleen sama, korvien kuulemasta ääni-informaatiosta on mahdollista päätellä tämä saapumisaikaero (TDOA, engl. *Time Difference Of Arrival*).

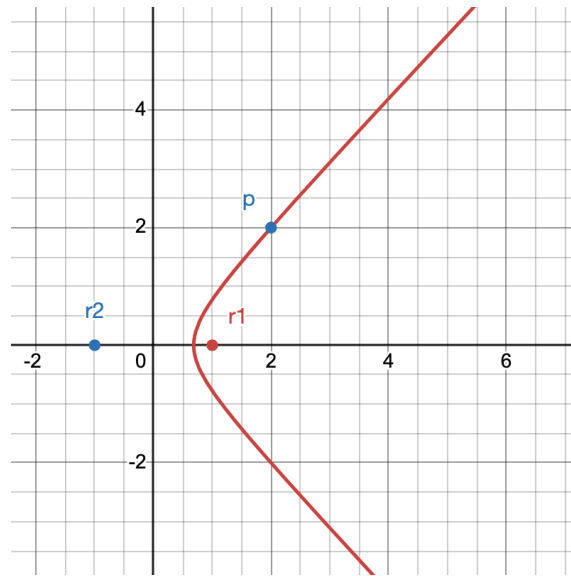


Kuva 2.1: Ääni etenee rajallisella nopeudella väliaineessa, joten eri etäisyydellä olevat mikrofonit $R1$ ja $R2$ vastaanottavat äänilähteestä $S1$ saapuvan signaalin eri aikaan ($t_1 - t_2$). Kuva tuotettu draw.io-sovelluksella (<https://app.diagrams.net/>).

Täsmälleen ottaen TDOA ei kahden vastaanottimen tapauksessa kuitenkaan itsenäisesti vielä riitä äänilähteen tarkkaan paikantamiseen. Yleisesti ottaen mielivaltaisesti valittua TDOA-arvoa vastaa ääretön määrä mahdollisia pisteitä, joista ääni on voinut saada alkunsa. Hahmotellaan karkeasti syytä tähän muodostamalla yhtälö kahden äänilähteen TDOA:sta kaksiulotteisessa avaruudessa, jossa väliaine on kaikkialla homogeenistä ja pysyy paikoillaan. Mielivaltainen äänilähde P , jolla on koordinaatit (x_0, y_0) , lähettää äänen, joka saapuu kahteen vastaanottimeen r_1 ja r_2 (joilla on vastaavasti koordinaatit (x_1, y_1) ja (x_2, y_2)) siten, että näiden TDOA on t , ja ääni saapuu ensin vastaanottimeen 1. Käytännössä voi ajatella TDOA:n olevan yhtä lailla etäisyysero, sillä voimme olettaa ääniaallon liikkuvan avaruudessa lineaarisesti, eli tietyssä ajassa yhtä pitkän matkan kaikkialla. Tällöin äänilähteen etäisyys vastaanottimeen 1 D_1 on

$$\sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2}.$$

Etäisyys vastaanottimeen 2 voidaan määrittää vastaavasti, mutta myös etäisyyden D_1 avulla: selvästi $D_2 = D_1 + t \iff t = D_2 - D_1$. Kaiken kaikkiaan yhtälön ratkaisut muodostavat kuvassa 2.2 näkyvän hyperbelikäyrän, eli kaikki tämän käyrän



Kuva 2.2: Esimerkki tilanteesta, jossa äänilähde sijaitsee pisteessä $(2, 2)$, vastaanotin r_1 pisteessä $(1, 0)$ ja vastaanotin r_2 pisteessä $(-1, 0)$. Käyrä vastaa niitä pisteitä, joissa sijaitseva äänilähde tuottaisi saman TDOA-arvon kuin $(2, 2)$. Kuva tuotettu Desmos-sovelluksella (<https://www.desmos.com/calculator>).

pisteet voisivat olla tietyn TDOA-arvon äänilähteen sijainteja. Vaikka esimerkissä tehdyt yksinkertaistavat oletukset pitäisivätkin paikkansa, on paikantamisessa siis hyödynnettävä muitakin keinoja. On myös merkillepantavaa, että äänilähteen etäisyyden kasvaessa suhteessa vastaanottimien keskinäiseen etäisyyteen yhtälön ratkaisut lähestyvät suoraa viivaa, joka kulkee vastaanottimien välisen keskipisteen kautta, koska vastaanottimien välisen etäisyyden merkitys pienenee äänilähteen loitontuessa niistä. Muun muassa näiden rajoitteiden aiheuttamiin ongelmiin paneudutaan seuraavassa luvussa.

3 Yleiskatsaus

paikannusmenetelmiin

Tässä luvussa luodaan pintapuolinen yleiskatsaus tärkeimpiin tekniikoihin, joita hyödynnetään modernissa äänilähteen paikannuksessa. On tärkeää huomata, että monet näistä tekniikoista eivät suinkaan ole toisiaan poissulkevia, ja monet tässä luvussa esiteltävät termit ovatkin enemmän kattonimityksiä joukolle tekniikoita, jotka hyödyntävät tiettyjä peruseriäiteitä. Lista teknologioista ei luonnollisesti myöskään ole millään tapaa tyhjentävä; kirjallisuudesta löytyy lukuisia muitakin teknologioita ja näiden variaatioita.

3.1 Paikannuksen lähtökohdat ja ongelmakuvaus

Ääniaallot ovat signaaleja, ja äänilähteen paikannus onkin erikoistus signaalinpaikannusongelmista. Monet perustavanlaatuiset tekniikat, kuten keilanmuodostus (engl. *beamforming*) tai TDOA, ovat yleistettävissä minkä tahansa ajassa ja paikassa etenevän signaalin, kuten esimerkiksi radioaallon, paikantamiseen. [3, p. 1] Koska ääniaallot etenevät verrattaen hitaasti ja väliaineessa, jonka ominaisuudet vaikuttavat merkittävästi aallon etenemiseen, täydellinen mallintaminen on usein hyvin hankalaa, raskasta tai jopa mahdotonta. Väliaineen liikkuminen esimerkiksi tuulen takia, ilmanpaineen vaihtelut tai ääniaallon siirtyminen väliaineesta toiseen ovat kaikki mallinnusta monimutkaistavia tekijöitä, ja useissa menetelmissä nämä te-

kijät oletetaan muuttumattomiksi vakioiksi; esimerkiksi, että ääniaalto liikkuu vain ilman läpi ilman väliaineen liikettä tai paineen ja lämpötilan muutoksia. Monissa sovelluksissa, esimerkiksi sisätiloissa, tämä oletus saattaa olla hyvinkin perusteltu. Tietysti jos käytetään sovellusta, jota kehitettäessä nämä oletukset on tehty, ja oletukset eivät käyttötapauksessa pidä, paikannustarkkuus saattaa muuttua merkittävästi heikommaksi tai jopa käytännössä olemattomaksi.

Äänilähteen paikannusta (SSL, engl. *Sound Source Localisation*) voi luonnehtia niin sanotuksi käänteiseksi tai inversio-ongelmaksi (engl. *inverse problem*). Inversio-ongelmissa jonkin alkuperäisen ongelman ratkaisu on tiedossa, ja on tarkoitus selvittää tämän alkuperäisen ongelman lähtöasetelma. Inversio-ongelmaan liittyvän alkuperäisen ongelman voi tässä tapauksessa määrittellä olevan ääniaallon propagoitumisen mallintaminen, kun äänilähteen ominaisuudet tunnetaan. Tällöin kaikki aaltojen lukemattomat ristivaikutukset, heijastumiset ja diffraktiot voidaan teoriassa aina laskea halutulla resoluutiolla. Voidaan ajatella, että resoluution loppuminen kesken ei ongelmana enää ole mallintamisen piirissä. Toiseen suuntaan tilanne on eri; kun nämä vaimentuneet ja lukuisaan otteeseen toistensa kanssa ristivaikuttaneet aallot alittavat mittausresoluution, muuttuu mahdottomaksi saada kaikkea informaatiota, joka tarvittaisiin alkuperäisen äänilähteen tarkkaan mallintamiseen. [3, p. 31]

Kauko- ja lähikenttämallinnus

Mallinnusperspektiivistä voi sanoa paikannusongelmien jakaantuvan kahteen pääluokkaan: lähikenttä- ja kaukokenttämallinnukseen (engl. *near-field* ja *far-field*). Lähikenttämallinnusta määrittää se, että äänilähteen etäisyys vastaanottimista suhteutettuna vastaanottajien etäisyyteen toisistaan on niin pieni, ettei äänen tulokulman voida käytännössä olettaa olevan sama kaikille vastaanottimille. Vastaavasti kaukokenttämallinnuksessa äänilähteen oletetaan olevan niin kaukana, että aaltorintama on käytännössä suora, ja näin ollen tulokulma kaikkiin vastaanottimiin sa-

ma.¹ [7, p. 17] Koska tässä tilanteessa oletetaan aaltorintaman kaarevuuden pysyvän samana ensimmäisestä havaintopisteestä viimeiseen (ts. aallon sisältämä energia ei leviä laajemmalle alueelle tällä välillä), tällöin myöskään äänen amplitudin ei oleteta ehtivän laskea matkalla ensimmäisestä havaintopisteestä viimeiseen. Usein äänilähde myös oletetaan pistemäiseksi; jos äänilähteen koko on suuri, myös tämä vaikuttaa siihen, voidaako kaukokenttämallinnusta käyttää [2, p. 191].

Kaukokenttämallinnuksen voi ajatella olevan helpompaa lähikenttämallinnukseen verrattuna, sillä äänilähteiden pistemäisyys, tulokulmien yhdensuuntaisuus ja aallon amplitudin vakioisuus yksinkertaistavat mallintamista. Tämä johtaa usein mallinnusalgoritmien aikakompleksisuuden pienenemiseen. Toisaalta yksinkertaistusten tekeminen mallintamisessa on yhtä lailla myös informaation hylkäämistä mallista. [7, p. 17] Autonominen robotti, jossa mikrofonit on aseteltava muutaman senttimetrin päähän toisistaan, ja jonka tehtävä on paikantaa kymmenien tai satojen metrien päässä olevia äänilähteitä, ja jossa laskentateho on kriittinen rajoite, ei sovelluksena välttämättä jätä varaa lähikenttämenetelmien käyttämiselle. Toisaalta luentosalissa, jonka katossa sijaitsevan mikrofoniverkon tehtävä on kyetä vahvistamaan mistä tahansa luentosalissa syntyvää ääntä, ei näitä rajoitteita ole. Tässä vertailuesimerkissä kannattaa kiinnittää erityishuomiota siihen, että kaukokenttämallinnuksessa tehtävien, mallinnusta yksinkertaistavien oletusten on oikeasti myös oltava perusteltuja, jotta mallinnus olisi luotettavaa; luultavasti tässä luentosaliesimerkissä näitä oletuksia rikottaisiin niin voimakkaasti, että kaukokenttämallinnuksella tehty sovellus ei toimisi ainakaan halutulla tavalla. Käytännössä onkin siis niin, että käyttötapaus pitkälti määrittää, minkä tyyppinen mallintaminen kuhunkin tilanteeseen sopii.

¹Tämä vastaa luvun 2 esimerkin tilannetta, jossa äänilähde on niin kaukana, että ratkaisujoukon muodostamaa hyperbeliä voidaan käsitellä suorana.

DOA sekä yleisesti termistöstä

Äänilähteen paikantaminen on laitteiston ja laskennan halventumisen myötä muuttunut nopeasti mielenkiintoiseksi tutkimusalaksi. Kun paikantamista oli mahdollista toteuttaa kaupallisesti järkevästi, sille löytyi nopeasti suuri määrä erilaisia sovelluksia robotiikasta, sotateollisuudesta, viihde-elektronikasta, telekonferenssiteknologiasta, kuulolaitteista, valvontajärjestelmistä ja monilta muilta aloilta [14]. Ongelman parissa on siis tehty paljon insinööriyötä, ja erilaisia teknologioita on lukuisia. Tästä johtuen alalla on paljon yleisessä käytössä olevia lyhenteitä, joita on pyritty luettelemaan termistösivulla.

DOA:lla viitataan yleisesti tulokulman määrittämiseen (engl. *Direction Of Arrival*) [7, p. 17]. DOA-termiä käytetään usein lähes synonyymisesti äänilähteen paikantamisen kanssa, koska monessa sovelluksessa riittää määrittää pelkkä tulokulma, ja etäisyyden määrittäminen on verrattaen hankalampaa [15]. Esimerkiksi kameran suuntauksessa, ääntä kohti liikkumisessa sekä luontevan virtuaalisen kuuloympäristön luomisessa riittää usein määrittää pelkkä tulokulma.

3.2 Aika-akseliin perustuvat tekniikat

TDOA

TDOA on yksi yleisimmin käytetyistä äänilähteen paikannusmenetelmistä. Se perustuu nimensä mukaisesti mikrofoni-verkon eri mikrofoniin saapumisaikaeroihin. Suuri osa varhaisesta tutkimuksesta DOA:n saralla perustui nimenomaan TDOA-pohjaisiin tekniikoihin tapauksissa, joissa äänilähteitä on kulloisellakin hetkellä vain yksi. Lukuisten yhtäaikaisten äänilähteiden paikantaminen on vaikeampi ongelma, ja sitä varten on kehitetty TDOA-pohjaisten tekniikoiden johdannaisia. [14]

GCC

Tähän asti käsitellyissä esimerkeissä on oletettu äänilähteen olevan joko “päällä” tai “pois päältä”, ja lisäksi tämän olevan yksikäsitteisesti määriteltävissä. Tosiasiallisesti kaikki vastaanottimet kuitenkin taltioivat kaiken ympäristön hälyäänen, eikä niillä ole lähtökohtaisesti kykyä päätellä pelkän signaalin avulla, mikä signaalin osakomponentti on sitä ääntä, joka halutaan paikallistaa. Useimmiten SSL-ongelmissa haluttu ääni kuitenkin eroaa ympäristön äänestä jotenkin, ja usein se on amplitudiltaan voimakkaampi.² GCC eli yleistetty ristikorrelaatio (engl. *Generalised Cross-correlation*) on tekniikka, jolla voidaan koettaa etsiä TDOA-arvoa. Ristikorrelaation yleinen määritelmä on

$$(f \star g)(\tau) = \int_{-\infty}^{\infty} \overline{f(t)} \cdot g(t + \tau) dt \quad ,$$

jossa $\overline{f(t)}$ on signaalifunktion $f(t)$ kompleksikonjugaatti. Ristikorrelaatiossa toisen signaalin voi ajatella “liukuvan” toisen yli viipeen τ vaihdellessa. Jos paikallistettava ääni on paljon voimakkaampi kuin taustahäly, funktion maksimiarvo saavutetaan, kun tämän paikallistettavan äänen aaltomuoto sovittuu eri vastaanottimien taltioimissa äänissä samaan kohtaan. Tämän maksimiarvon antama viipeen τ arvo on siten TDOA-arvio. Tässä lähestymistavassa on tosin ongelmansa. Signaaleissa saattaa esiintyä voimakkaasti vaiheista, ei-toivottua ääntä, ja hälyäänen taso saattaa olla korkea. Yleistetty ristikorrelaatio vastaa näihin ongelmiin ikkunoimalla signaalin tietyssä vaiheessa prosessointia. [5]

²Vastaanottimen signaaliketjuun voidaan sisällyttää niin sanottu *portti* (engl. *gate*), joka laukeaa vain signaalin amplitudin ylittäessä tietyn arvon. Näin voidaan yrittää päätellä, onko paikallistettava ääni juuri saapunut vastaanottimeen.

GCC-PHAT

GCC-PHAT eli yleistetty ristikorrelaatio vaihemuunnoksella (engl. *Generalised Cross-correlation Phase Transform*) on äänilähteen paikannusmenetelmä, joka perustuu edellä mainittuun yleistettyyn ristikorrelaatioon sekä vaihemuunnokseen. Vaihemuunnoksella signaalit normalisoidaan taajuusavaruudessa. Tämä tehostaa menetelmää tilanteissa, joissa äänisignaali on kohinaisessa ympäristössä. [5]

3.3 Spektrianalyysiin perustuvat tekniikat

MUSIC

MUSIC, eli monisignaali luokittelu (engl. *Multiple Signal Classification*), on R. Schmidtin vuonna 1986 esittelemä spektripohjainen äänilähteen paikannusmenetelmä, joka perustuu mikrofoni-verkon signaalien taajuusspektrien analyysiin [6]. Tämä menetelmä hyödyntää signaalin aliavaruuksien (engl. *Subspace*) etsintää ja luokittelee äänilähteitä niiden taajuuskomponenttien perusteella. MUSIC on nimensä mukaisesti erikseen kehitetty sellaisia ympäristöjä varten, joissa on lukuisia äänilähteitä, kuten konserttisaleissa tai konferenssituloissa, ja se onkin yksi ensimmäisistä laajalti käytetyistä paikannusalgoritmeista, jotka kykenevät yhtäaikaaisesti useamman äänilähteen paikantamiseen. [14]

SRP

SRP, eli suunnatun vastevoiman analyysi (engl. *Steered Response Power*), on menetelmä, jossa äänilähdettä etsitään keilanmuodostuksella ruudukoksi jaotellusta avaruudesta. Näistä hypoteettisista äänilähteiden sijainneista se, joka todennäköisiten tuottaisi havaitun signaalin, tulkitaan äänilähteen sijainniksi. [8] Keilanmuodostuksessa taas sovitetaan vastaanottimien keskinäiset viipeet sellaisiksi, että tietystä

suunnasta saapuva signaali taltioituu kaikkiin vastaanottimiin siten, että se interferoi itsensä kanssa konstruktivisesti, eli eri vastaanottimien signaalien huiput ovat kohdakkain kyseiselle signaalille.³ [7]

SRP-PHAT

SRP-PHAT, eli suunnatun vastevoiman analyysi vaihemuunnoksella (engl. *Steered Response Power Phase Transform*), on kehittynyt äänilähteen paikannusmenetelmä, joka yhdistää suunnatun vastevoiman analyysin ja vaihemuunnoksen. Vaihemuunnosta hyödynnetään tässä teknologiassa samoin kuin GCC-PHAT:issa. Tällä menetelmällä saavutetaan usein korkea tarkkuus äänilähteen paikannuksessa hankalissa ympäristöissä, kuten kaikusissa huoneissa tai ulkotiloissa. [8]

³Ristikorrelaatiota voisi ajatella keilanmuodostuksen inversio-ongelmana: mikä viive aiheuttaisi konstruktivisen interferenssin kahdelle signaalille?

4 Menetelmien vertailua

Tässä luvussa paneudutaan äänenpaikannustutkimuksen nykytilaan. Kenttää määrittävät erityisesti käytännön sovellusten määrittämien resurssirajoitteiden puitteisissa toimiminen, ja eri menetelmissä onkin merkittäviä eroja käyttökelpoisuudessaan äänenpaikannussovelluksiin. Luvussa tarkastellaan myös hiljattain kehitettyjä uusia muunnelmia luvussa 3 esitellyistä menetelmistä sekä vertaillaan näillä saavutettuja tuloksia.

4.1 Laskentakustannukset ja IoT

IoT eli asioiden internet (engl. *Internet of Things*) on etenkin 5G-ajan alettua kasvattanut suosiotaan voimakkaasti, ja uusia IoT-sovelluksia on kehitetty paljon. Useimmissa IoT-sovelluksissa laitteiston valmistuskustannukset ja energiankulutus on pidettävä matalina. Nämä rajoitteet määrittävät sovellussuunnittelua ja toimivatkin luonnollisena kannustimena SSL-algoritmien optimoinnille. Toisaalta monet ympäristöä havainnoivat IoT-laitteet on jo lähtökohtaisesti varustettu sensoreilla kuten SSL:ssä tarvittavilla mikrofoneilla, joten laitteistoympäristö paikannussovelluksille on usein jo määritelmällisestikin käytössä. On huomattava, että sovelluksissa, joissa on lukuisia antureita sekä näihin yhteydessä oleva isäntäyksikkö, voitaisiin äänitetyt signaalit teoriassa lähettää isäntäyksikön prosessoitavaksi. Monet IoT-sovellukset ovat kuitenkin akkukäyttöisiä, ja signaalin lähettäminen verkon yli riittävän korkearesoluutioisena pienentäisi sensorien akunkestoja niin paljon, että käytännössä

signaalinprosessointia kannattaa usein suorittaa jo sensoreissa itsessään. [16] Esimerkiksi Latif ym. [17] esittelevät konseptin hankaliin pelastustehtäviin soveltuvi- ta, äänenpaikannukseen kykenevistä bioboteista, joissa *Gromphadorhina portentosa* -lajin torakoihin integroidaan verkkoon kytketty siru. Tässä prototyypissä tosin signaali lähetetään isäntäyksikköön, jossa DOA-estimaatio suoritettiin MATLAB-sovelluksessa. Joka tapauksessa tällaisen sovelluksen suunnittelussa on selvää, että akunkesto muodostuu ratkaisevaksi ongelmaksi. Laskenta-ajan näkökulmasta kuitenkin ilmenee, että esimerkiksi SRP-pohjaiset tekniikat voivat hyödyntää rinnak- kaisuutta, koska tehokartan (engl. *power map*) yksittäisten ehdokaspisteiden tehoar- vojen laskennat eivät riipu toisistaan ja ovat näin rinnakkaistettavissa [16].

Käytännössä korkeammalla laskenta-ajalla saavutetaan korkeamman resoluution paikannus, ja vastavuoroisesti korkeampi paikannusresoluutio nostaa laskenta-aikaa. Koska hyvin monessa sovelluksessa on vaatimus reaaliaikaisuudesta, suuri osa alalla tapahtuvasta optimointitutkimuksesta on keskittynyt juuri madaltamaan laskenta- kustannuksia. [16] Toisaalta on myös monia kehityksiä, joissa on keskitytty nimenom- aan menetelmien tarkkuuden parantamiseen. Nunes ym. [18] esimerkiksi kertovat saavuttaneensa kokoussalissa n. 7 senttimetrin tarkkuuden paikannuksessa kehittä- mällään SRP:hen pohjautuvalla hierarkkisen etsinnän algoritmilla. Tutkimuksessa tosin mainitaan metodin myös laskevan laskentavaatimuksia merkittävästi. Myös Keyrouz ym. [19] nostavat esiin korkean paikannustarkkuuden humanoidiroboteille suunnatussa, binauraalisessa paikannusjärjestelmässään. Tässä on huomattava, että kyseisen prototyypin koearjoissa ei tunnu olevan ainakaan eksplisiittisesti huo- mioitu matalan energiankulutuksen asettamia vaatimuksia laskentateholle. Tämä saattaa osaltaan indikoida, että alalla tapahtuva kehitys, joka tähtää nimenomai- sesti tarkempaan paikannukseen, ei välttämättä ole suuntautunut IoT-kehitykseen. Toki on myös paljon sovelluskohteita, joissa signaali voidaan suoraan lähettää te-

hokkaaseen keskustietokoneeseen ilman huolta akunkestosta, kuten vaikkapa luentosaliesimerkissä; käytännössä kaikki kiinteät asennukset lukeutuvat tähän luokkaan.

4.2 Sovellusympäristön tuomat haasteet

Äänilähteen paikannusmenetelmien suorituskyky vaihtelee merkittävästi ympäristön rajoitteiden kuten tilan kaikuisuuden ja korkean hälytason mukaan. Puhtaasti TDOA:han ja GCC:hen perustuvat menetelmät vaativat verrattaen pienen määrän numeerisia operaatioita verrattuna SRP-pohjaisiin menetelmiin. Nämä tekniikat ovat kuitenkin herkkiä ympäristön signaali-hälysuhteelle (SNR, engl. *Signal-to-noise Ratio*) sekä tilan kaikuisuudelle. [18] Tämä johtuu siitä, että voimakkaat kaiut ja korkea hälytaso vähentävät ristikorrelaation maksimiarvon yksikäsitteisyyttä; funktion ”piikki” ei siis ole yhtä huipukas tällaisissa ympäristöissä.

SRP-pohjaiset menetelmät ovat huomattavasti robustimpia matalalle SNR:lle.¹ SRP:n korkeammat laskentavaatimukset asettavat kuitenkin optimointipaineita SRP-pohjaisille metodeille. Nunesin ym. [18] tapaan myös Yook ym. [20] lähestyvät optimointiongelmia SRP:ssä tapahtuvan ruudukkoetsinnän (engl. *grid search*) optimoinnin kautta käyttämällä kaksitasoista hakuavaruuden klusterointia ja raportoivat saavuttaneensa 61,8 prosentin laskun laskentakustannuksissa verrattuna naiviin ruudukkoetsintään. [20]

Signaaliavaruuksiin perustuvat menetelmät, kuten MUSIC, vaativat lukuisia mikrofoneja toimiakseen tarkasti. Lisäksi, kuten keilanmuodostukseen perustuvat SRP-pohjaiset menetelmät, nämäkin ovat laskennallisesti raskaampia menetelmiä verrattuna TDOA-pohjaisiin menetelmiin. [15] Lisäksi esimerkiksi MUSIC:in tarkkuus heikkenee kaikuisissa ympäristöissä merkittävästi [21], mikä saattaa olla yhteydessä siihen, että MUSIC on nimensä mukaisesti suunniteltu useamman ääniläh-

¹Matala SNR tarkoittaa signaalin olevan heikko suhteessa hälyyn, eli voimakasti hälyistä ympäristöä.

teen havaitsemiseen, ja kaikuisissa ympäristöissä äänilähteen jälkikaiku käyttäytyy monilta osin samoin kuin erillinen äänilähde käyttäytyisi.

Sovellusympäristössä saattaa olla myös piirteitä, joita voidaan hyväksikäyttää menetelmien suunnittelussa. Wang ym. [15] hyödyntävät mobiilille robotille suunnitellussa äänenpaikannusmenetelmässään robotin liikettä; liikkeen tuoma lisäinformaatio helpottaa äänilähteiden sijainnin päättelyä, koska robotin liikkumishistoria 3D-ympäristössä on systeemin tiedossa.

4.3 Koneoppimisen vaikutus

Laskentatehon halventuminen on mahdollistanut hermoverkkojen tehokkaamman hyödyntämisen monella alalla, ja tämä heijastuu myös SSL-ongelmiin. Ensimmäiset käyttökelpoiset, perinteiseen MLP-verkkoon (engl. sanoista *Multi-layer Perceptron*) perustuvat paikannusmenetelmät esiteltiin 2015 ja 2016 [22]. Hermoverkkojen hyödyntäminen äänen paikannuksessa tuo kuitenkin mukanaan erillisongelman: verkko on koulutettava oikeita paikannustilanteita edustavalla datalla. Diaz-Guerra ym. [22] esittävät menetelmän, jossa SRP-PHAT-tekniikalla lasketut tehokartat (engl. *power map*) toimivat syötteenä kolmiulotteiselle konvoluutionaaliselle hermoverkolle (CNN, engl. *Convolutional Neural Network*). Verkko on koulutettu hyödyntäen sekä akustisia simulaatioita että LOCATA-aineistoa [23], joka koostuu oikeista lukuisilla mikrofoneilla äänitetyistä äänitteistä sekä näiden äänitteiden sijaintitiedoista. Myös Lee ym. [24] esittelevät menetelmän, joka hyödyntää CNN-verkkoja paikallistamisessa. Tässä menetelmässä hermoverkon rooli on vähemmän suoraviivainen, mutta siinäkin hyödynnetään SRP:n muodostamaa tehokarttaa, jotta löydettäisiin piirteitä, joita voidaan hyödyntää paikallistamisessa [24].

Monet CNN-pohjaiset paikannusmenetelmät ovat kuitenkin *ei-kausaalisia* aika-akselilla, eli ne hyödyntävät kulloisellakin ajanhetkellä informaatiota tulevaisuudesta. Tämä tekee näistä menetelmistä hyvin hankalia tai jopa mahdottomia toteuttaa

reaaliaikaisesti. [22] Diaz-Guerran ym. esittelemä menetelmä on kuitenkin toteutettu täysin kausaalisesti, mikä tekee siitä reaaliaikaisiin sovelluksiin kelpaavan [22], ja näin ollen validoi CNN-pohjaisten menetelmien käyttökelpoisuuden reaaliaikaisissa sovelluksissa.

5 Yhteenveto

Tässä tutkielmassa selvitettiin äänilähteen paikantamista insinööriongelmana. Ensin perehdyttiin lyhyesti fysikaaliseen ja matemaattiseen viitekehykseen, jossa äänenpaikannus mikrofoni-verkon avulla tapahtuu. Sitten luotiin yleiskatsaus menetelmiin ja niiden toimintaperiaatteisiin. Lopuksi tarkasteltiin äänenpaikannustutkimuksen nykytilaa ja sitä, miten esiteltyjä menetelmiä on kehitetty eteenpäin ja sovellettu reaali maailman ongelmiin.

Äänilähteen paikannus olisi matemaattisia työkaluja hyödyntäen helposti ratkaistavissa oleva ongelma, jos käytännön sovelluksissa voisi olettaa äänilähteen olevan helposti tunnistettavissa. Osoittautuu kuitenkin, että näin ei ole, ja menetelmien toiminnasta suuri osa on sitä, että signaalista koetetaan selvittää, mikä sen osakomponentti on kiinnostuksen ja näin myös paikannuksen kohteena olevaa ääntä, ja mikä on kohinaa.

Kehitettyjä paikannusmenetelmiä on lukuisia, ja niiden kaikkien käsittelyyn ei luonnollisesti tämän tutkielman viitekehyksessä ollut aikaa. Havaittiin kuitenkin, että äänilähteen etäisyyden paikantaminen on hankalaa, ja suuri osa sovelluksista keskittyy vain tulokulman määrittämiseen. Tärkeimmät teknologiat, jotka ovat perinteisessä mikrofoni-verkkopaikallistamisessa käytössä, lienevät aika-akseliin perustuvat tekniikat, kuten TDOA ja GCC, sekä spektrianalyysiin perustuvat, kuten SRP-PHAT. Lisäksi signaalialivaruuksiin perustuvilla menetelmillä, kuten MUSIC:illa, on omat sovelluskohteensa.

Alan nykytutkimusta määrittävät voimakkaasti IoT-alan nopea kasvu, sovellusympäristön tuomat haasteet sekä koneoppimismenetelmien tehokkaan hyödyntämisen mahdollisuus. IoT:n vaikutuksesta sovelluskehityksen painopiste on usein laskenta-ajan ja rinnakkaisuuden optimoinnissa. Samoin lukuisat uudet IoT:n mahdollistamat sovellusympäristöt, kuten meluisat ulkotilat ja sensorien liikkuvuus, vaikuttavat voimakkaasti menetelmien suosioon. SRP-PHAT nousee esiin useassa sovelluksessa robustina ja tarkkana menetelmänä, jossa on verrattaen korkeista laskentavaatimuksistaan huolimatta lukuisia optimoitavissa olevia osa-alueita. Myös muilla käsitellyillä teknologioilla on yhä paikkansa laajalla ja monimuotoisella sovelluskentällä.

Äänilähteen paikantamisella on suuri määrä mahdollisia sovelluksia, rakennussuunnittelusta puolustusteollisuuteen ja lääketieteeseen. Se ei kuitenkaan laskennallisena ongelmana ole triviaali, ja käytännön hankalat sovellusympäristöt, kuten meluisat tai kaikuisat tilat, hankaloittavat paikantamista entisestään. Paikannusteknologiat ovat kuitenkin onnistuneet vastaamaan näihin haasteisiin, ja laskentatehon halventuminen ja IoT:n kasvu on mahdollistanut suuren määrän uutta tutkimustyötä ongelman parissa. Myös koneoppimismenetelmät ovat fuusioituneet olemassa oleviin paikannusteknologioihin, ja edistyneisimmät menetelmät ovatkin usein hyvin pitkälle optimoituja komposiittimenetelmiä.

Koneoppimisen tämänhetkisen nopean kehityksen voisi olettaa tuovan nämä menetelmät vielä olennaisemmaksi osaksi äänilähteen paikantamisteknologiaa, ja tulemme varmasti näkemään lähitulevaisuudessa yhä tarkempia ja tehokkaampia äänilähteen paikannusmenetelmiä.

Lähdeluettelo

- [1] D. Self, *Acoustic Location and Sound Mirrors*, <https://web.archive.org/web/20110112224410/http://www.aqpl43.dsl.pipex.com/MUSEUM/COMMS/ear/ear.htm>, Arkistoitu alkuperäisestä 12.1.2011. Haettu 20.5.2024.
- [2] L. E. Kinsler, *Fundamentals of acoustics*, L. E. Kinsler, toim. John Wiley & Sons, 1999, ISBN: 1-5231-1788-5.
- [3] M. Bai, *Acoustic array systems : theory, implementation, and application*, J. Benesty ja J.-G. Ih, toim. John Wiley & Sons, Inc., 2013, ISBN: 1-299-31596-8. DOI: 10.1002/9780470827253.
- [4] C. Shannon, "Communication In The Presence Of Noise", eng, *Proceedings of the IEEE*, vol. 86, nro 2, s. 447–457, 1998, ISSN: 0018-9219.
- [5] C. Knapp ja G. Carter, "The generalized correlation method for estimation of time delay", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, nro 4, s. 320–327, 1976. DOI: 10.1109/TASSP.1976.1162830.
- [6] R. Schmidt, "Multiple emitter location and signal parameter estimation", *IEEE Transactions on Antennas and Propagation*, vol. 34, nro 3, s. 276–280, 1986. DOI: 10.1109/TAP.1986.1143830.
- [7] J. H. DiBiase, "A high-accuracy, low-latency technique for talker localization in reverberant environments using microphone arrays", tohtorinväitös-kirja, Brown University, 2000.

- [8] J. H. DiBiase, H. F. Silverman ja M. S. Brandstein, ”Robust Localization in Reverberant Rooms”, teoksessa *Microphone Arrays: Signal Processing Techniques and Applications*, M. Brandstein ja D. Ward, toim. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001, s. 157–180, ISBN: 978-3-662-04619-7. DOI: 10.1007/978-3-662-04619-7_8.
- [9] Tampereen yliopiston Musiikintutkimuksen laitos, *mute: musiikin teoriaa webissä*, <https://web.archive.org/web/20170218025707/http://www15.uta.fi/arkisto/mustut/mute/aai01.htm>, Arkistoitu alkuperäisestä 18.2.2017. Haettu 21.5.2024, 1998.
- [10] Sony ja Philips, *Compact Disc Digital Audio System*, IEC 60908:1987, also known as the Red Book, International Electrotechnical Commission (IEC), Geneva, Switzerland, 1980.
- [11] C. J. Plack, *The Sense of Hearing*. Psychology Press Ltd, 2014, ISBN: 978-1848725157.
- [12] Diffen LLC, *Mono vs. Stereo*, https://web.archive.org/web/20230117071148/https://www.diffen.com/difference/Mono_vs_Stereo, Arkistoitu alkuperäisestä 17.1.2023. Haettu 21.5.2024.
- [13] National Weather Service, *Speed of Sound Calculator*, https://www.weather.gov/epz/wxcalc_speedofsound, Haettu 21.5.2024.
- [14] D. Pavlidi, A. Griffin, M. Puigt ja A. Mouchtaris, ”Real-time multiple sound source localization and counting using a circular microphone array”, *IEEE Transactions on Audio, Speech and Language Processing*, vol. 21, s. 2193–2206, 10 2013, ISSN: 15587916. DOI: 10.1109/TASL.2013.2272524.
- [15] Z. Wang, W. Zou, H. Su, Y. Guo ja D. Li, ”Multiple Sound Source Localization Exploiting Robot Motion and Approaching Control”, *IEEE Transactions on*

- Instrumentation and Measurement*, vol. 72, 2023, ISSN: 15579662. DOI: 10.1109/TIM.2023.3298406.
- [16] J. A. Belloch, J. M. Badia, F. D. Igual ja M. Cobos, ”Practical considerations for acoustic source localization in the IoT era: Platforms, energy efficiency, and performance”, *IEEE Internet of Things Journal*, vol. 6, s. 5068–5079, 3 kesäkuu 2019, ISSN: 23274662. DOI: 10.1109/JIOT.2019.2895742.
- [17] T. Latif, E. Whitmire, T. Novak ja A. Bozkurt, ”Sound Localization Sensors for Search and Rescue Biobots”, *IEEE Sensors Journal*, vol. 16, s. 3444–3453, 10 toukokuu 2016, ISSN: 1530437X. DOI: 10.1109/JSEN.2015.2477443.
- [18] L. O. Nunes, W. A. Martins, M. V. Lima et al., ”A steered-response power algorithm employing hierarchical search for acoustic source localization using microphone arrays”, *IEEE Transactions on Signal Processing*, vol. 62, s. 5171–5183, 19 lokakuu 2014, ISSN: 1053587X. DOI: 10.1109/TSP.2014.2336636.
- [19] F. Keyrouz, ”Advanced binaural sound localization in 3-D for humanoid robots”, *IEEE Transactions on Instrumentation and Measurement*, vol. 63, s. 2098–2107, 9 2014, ISSN: 00189456. DOI: 10.1109/TIM.2014.2308051.
- [20] D. Yook, T. Lee ja Y. Cho, ”Fast Sound Source Localization Using Two-Level Search Space Clustering”, *IEEE Transactions on Cybernetics*, vol. 46, s. 20–26, 1 tammikuu 2016, ISSN: 21682267. DOI: 10.1109/TCYB.2015.2391252.
- [21] N. Stefanakis, D. Pavlidi ja A. Mouchtaris, ”Perpendicular Cross-Spectra Fusion for Sound Source Localization with a Planar Microphone Array”, *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 25, s. 1821–1835, 9 syyskuu 2017, ISSN: 23299304. DOI: 10.1109/TASLP.2017.2718733.
- [22] D. Diaz-Guerra, A. Miguel ja J. R. Beltran, ”Robust Sound Source Tracking Using SRP-PHAT and 3D Convolutional Neural Networks”, *IEEE/ACM Tran-*

-
- sactions on Audio Speech and Language Processing*, vol. 29, s. 300–311, 2021, ISSN: 23299304. DOI: 10.1109/TASLP.2020.3040031.
- [23] H. W. Löllmann et al., ”The LOCATA Challenge Data Corpus for Acoustic Source Localization and Tracking”, teoksessa *Proceedings of the IEEE 10th Sensor Array and Multichannel Signal Processing Workshop*, heinäkuu 2018, s. 410–414.
- [24] S. Y. Lee, J. Chang ja S. Lee, ”Deep Learning-Enabled High-Resolution and Fast Sound Source Localization in Spherical Microphone Array System”, *IEEE Transactions on Instrumentation and Measurement*, vol. 71, 2022, ISSN: 15579662. DOI: 10.1109/TIM.2022.3161693.