Aki Kärkkäinen

SPATIAALINEN ÄÄNI KÄYTTÖLIITTYMÄSSÄ PIILOINFORMAATION VÄLITTÄJÄNÄ JA VISUAALISEN KUORMAN KEVENTÄJÄNÄ

Tietojärjestelmätieteen
pro gradu -tutkielma
12.6.2002

Jyväskylän yliopisto
Tietojenkäsittelytieteiden laitos
Jyväskylä

TIIVISTELMÄ

Kärkkäinen, Aki Antero
Spatiaalinen ääni käyttöliittymässä piiloinformaation välittäjänä ja visuaalisen kuorman keventäjänä / Aki Kärkkäinen
Jyväskylä: Jyväskylän yliopisto, 2002.
82 s.
pro gradu -tutkielma

Tutkimuksessa tarkastellaan spatiaalista ääntä graafisen käyttöliittymän laajentajana. Tavoitteena on tutkia, miten ääntä voitaisiin käyttää grafiikan kanssa piiloinformaation välittämisessä. Käsitteellis-teoreettisen analyysin avulla pyritään sijoittamaan keskeiset käsitteet laajempaan viitekehykseen.

Tutkimuksessa tarkastellaan ensin kuuntelemiseen ja äänen paikantamiseen liittyviä seikkoja, jonka jälkeen esitetään synteettisen spatialisoinnin periaate ja sen toteutukseen liittyvät ongelmat. Tämän jälkeen spatiaalista ääntä tarkastellaan käyttöliittymäkontekstissa semioottisessa viitekehyksessä.

Tulokset osoittavat, että spatiaalisia - syntaksia noudattavia - jokapäiväisiä äänitapahtumia voidaan liittää käyttöliittymän tehtäviin ja objekteihin paljastamalla siten muutoin piiloon jäävää informaatiota. Jatkotutkimusaiheita on runsaasti psykoakustiikan, käytännön sovellusten ja akustisen ekologian piirissä. Erityisesti liikkuvan tietojenkäsittelyn pienikokoisissa laitteissa äänen merkitys korostunee tulevaisuudessa, visuaalisen informaation määrän supistuessa.

AVAINSANAT: akustinen ekologia, auralisaatio, käyttöliittymä, piiloinformaatio, psykoakustiikka, spatialisointi, äänen paikantaminen, ikoninen ääni

KIITOKSET

Kiitän seuraavia henkilöitä ja laitoksia siitä inspiraatiosta, jota ilman tämä työ ei olisi syntynyt:

Tauri Kankaanpää, Laëtitia Kulyk, Gilles Martin, Juha Oravala, Bettina Ortmann, Marinê Pereira, Seppo Puuronen ja Kai Tuuri (työn kaksi tarkastajaa), Leena Uski, Panu Varstala, Henna Välikangas, Äiti ja Isä, Jyväskylän yliopisto sekä seuraavat laitokset Pariisissa: Centre Pompidou, Cité des Sciences et de l'Industrie (La Villette), IRCAM (Institut de Recherche et Coordination Acoustique/Musique) sekä Université de la Sorbonne nouvelle Paris III (Cinéma et Audiovisuel).

SISÄLLYSLUETTELO

1 JOHDANTO 1

1.1 TUTKIMUKSEN TAUSTAA 1

1.2 TUTKIMUSTEHTÄVÄ 5

1.3 SOVELLETTAVAT TIETEENALAT JA TUTKIMUKSEN RAKENNE 8

2 KUUNTELEMALLA TIETOA MAAILMASTA 10

2.1 ÄÄNI JA SEN ETENEMINEN KORVAAN 10

2.2 ÄÄNTEN RYHMITTELY 12

2.3 AISTIHAVAINTOJÄRJESTELMÄT 14

2.4 AUDIOVISUAALINEN SOPIMUS 15

2.5 JOKAPÄIVÄINEN JA MUSIIKILLINEN KUUNTELU 19

2.6 SYNTAKTISET JA SEMANTTISET TEKIJÄT ÄÄNTEN TUNNISTAMISESSA 22

3 SPATIAALINEN ÄÄNI 26

3.1 ÄÄNEN PAIKANTAMINEN 27

3.2 SIIRTOFUNKTIO 30

3.3 ÄÄNEN SYNTEETTINEN SPATIALISOINTI 31

3.4 AURALISAATIO 35

3.5 KAIUTIN- JA KUULOKEÄÄNENTOISTON VASTAKKAINASETTELU 36

3.6 ÄÄNEN SPATIALISOINNIN ONGELMAT 37

3.7 SPATIAALISTA ÄÄNTÄ HYÖDYNTÄVIÄ SOVELLUKSIA 39

4 ÄÄNIMAISEMA-ANALYYSI KÄYTTÖLIITTYMÄSSÄ 43

4.1 HILJAISUUS 43

4.2 ÄÄNEN KÄYTTÖTARKOITUS 45

4.3 IKONISET ÄÄNET 49

4.4 SYMBOLISET ÄÄNET 53

4.5 METAFORISET ÄÄNET 55

4.6 ÄÄNITAPAHTUMIEN LUOKITTELUJÄRJESTELMÄT 59

5 KESKUSTELUA 64

5.1 TULOKSET 64

5.2 JOHTOPÄÄTÖKSET 68

6 YHTEENVETO 71

LÄHTEET 73

KUVIOT

KUVIO 1. OBJEKTIN ÄÄNELLINEN JA VISUAALINEN ESITYS 20

KUVIO 2. BINAURAALINEN ÄÄNITYS JA TOISTO 32

KUVIO 3. SIIRTOFUNKTIOIDEN MITTAUS JA SIMULOINTI 33

TAULUKOT

TAULUKKO 1. ÄÄNITUTKIMUKSEN KATEGORIAT 6

TAULUKKO 2. TUTKIMUKSEN TULOKSET 70

1 JOHDANTO

Luvussa esitellään ensin aihepiiri ja sen rajaukset, jonka jälkeen määritellään tutkimusongelma sekä eri tieteenalojen osuus tutkimuksessa.

1.1 Tutkimuksen taustaa

Miksi ääntä pitäisi käyttää käyttöliittymässä? Onko ääni vain turha ja irrallinen elementti, joka häiritsee keskittymistä tehtävään? Tällaisia kysymyksiä esitetään yleensä silloin, kun äänisuunnittelu on tehty puutteellisesti tai kun sitä ei osata tehdä. Vaikka kuulolla koettu ilmiö edustaa fenomenologisesti yhtä varmaa tietoa kuin näöllä koettu, nähtyä pidetään objektiivisena ja kuultua subjektiivisena tietona (Jauhiainen 1995, 13). Äänen hyödyntäminen graafisissa käyttöliittymissä onkin toistaiseksi jäänyt todella vähäiseksi. Grafiikkaa ja ääntä on kuitenkin käytännössä vaikea erottaa toisistaan. Niiden tulisikin tukea ja täydentää toisiaan siten, ettei kumpikaan ole täydellinen ilman toista.

Mikä sitten voisi olla äänen tuottama lisäarvo graafisessa käyttöliittymässä? Chionin (1990, 5) mukaan ääni rikastuttaa kuvaa antamalla vaikutelman, että ääni on "turha", samaan aikaan kun ääni tuo esille ja voimistaa sen, mitä kuvassa jo on (mutta joka ei tulisi esille ilman ääntä!). Vaikka Chion puhuu yksinomaan elokuvasta, näkökulma on silti mielenkiintoinen ja sovellettavissa käyttöliittymään: se sisältää ajatuksen, että ääni voi olla informatiivinen ja ekspressiivinen olematta silti häiritsevä. Ääni on kuulunut elokuvaan jo 70 vuotta (Chion 1990, 142), miksei tämä kehitys kuuluisi myös käyttöliittymissä?

Ääntä on perinteisesti käytetty käyttöliittymässä varoitus- tai ilmoitusmerkityksessä tai tunnelman luomisessa (lähinnä pelit ja multimediasovellukset). Äänen tarkoituksena on ollut myös auttaa näkövammaisia työskentelemään tietokoneen kanssa. Ajatus on luonnollinen sikäli, että käytetäänhän grafiikkaakin vain näkeviä ihmisiä varten. Tällainen näkökulma on kuitenkin liian rajoittunut. Tuntuu keinotekoiselta ajatella, että näkevät ihmiset käyttäisivät graafista käyttöliittymää ja näkövammaiset kuulokäyttöliittymää. Käsite kuulokäyttöliittymä on - toisin kuin graafinen tai merkkipohjainen käyttöliittymä - vielä verraten harvinainen.

Kuulokäyttöliittymätutkimus on perinteisesti jakautunut kahteen tutkimushaaraan: graafisen käyttöliittymän laajentamiseen äänen avulla (auditory interface, auditory display) ja näkövammaisille tarkoitettuun, graafisen käyttöliittymän korvaavaan kuulokäyttöliittymään (audio-only interface, nonvisual interface). Käsitteet menevät osittain päällekkäin; joskus auditory interface saattaa viitata myös näkövammaisille tarkoitettuun käyttöliittymään (ks. Edwards 1988). Pääasia on, että käsitteillä viitataan informaation esittämiseen käyttöliittymässä äänen avulla.

Ensin täytyy kuitenkin selvittää, minkälaisia ääniä käytetään. Tässä tutkimuksessa keskitytään synteettisiin tai luonnollisiin ei-puheääniin, koska puheääni on (a) hidasta, (b) ärsyttävää kuunnella ja (c) liian huomiota vaativaa. Informaatiota voi jäädä piiloon (hidden information) esimerkiksi siksi, että näyttöön on ahdettu liikaa visuaalista informaatiota (Blattner, Papp III & Glinert 1994, 448). Cohenin (1994, 500) mielestä tietokoneen näyttö ei yksinkertaisesti ole tarpeeksi suuri kyetäkseen näyttämään visuaalisesti kaikki käynnissä olevat toiminnot. Suuri osa näistä toiminnoista tapahtuu käyttäjä "selän takana", oli sitten kyseessä käyttäjän alullepanema toiminto tai jokin käyttäjää kohti suuntautuva toiminto. (Cohen 1994, 500.)

Ensiksi mainittu tutkimushaara (graafisen käyttöliittymän laajentaminen) jakaantuu kolmeen pääsuuntaukseen: *ikonisiin ääniin* (auditory icons), *symbolisiin ääniin* (earcons) ja *sonifikaatioon* (sonification). Ikoniset äänet perustuvat jokapäiväisen kuuntelun tuottamaan informaatioon (Gaver 1986, 168). Symboliset äänet ovat abstrakteja, synteettisiä ääniä, joita voidaan käyttää strukturoidusti äänellisten viestien esittämiseen käyttöliittymässä (Brewster, Wright & Edwards 1993, 222). Sonifikaatio tarkoittaa ohjelmien tuottaman tiedon esittämistä äänen avulla (Hereford & Winn 1994, 219).

Herefordin ja Winnin (1994, 212-213) mukaan äänellä on käyttöliittymässä kaksi funktiota: ääni (a) kertoo käyttäjälle järjestelmän tilasta kullakin hetkellä hälytysäänillä tai "tarkkailee" taustalla kunnes käyttäjä vaatii tietoa järjestelmän tilasta ja (b) välittää ohjelmien tuottamaa tietoa. Hereford ja Winn (1994, 213) luokittelevat ikoniset ja symboliset äänet ryhmään (a) ja sonifikaation ryhmään (b). Sonifikaatio sivuutetaan tässä tutkimuksessa. Jälkimmäinen tutkimushaara (näkövammaisille tarkoitetun käyttöliittymän tutkimus) rajataan myöskin tämän tutkimuksen ulkopuolelle. Aihepiiriä ovat lähestyneet ainakin Edwards (1988), Mynatt ja Weber (1994) sekä Mereu ja Kazman (1996).

Ihmisen ja koneen vuorovaikutusta tarkasteltaessa tulisi käyttää sateenkaarikäsitettä käyttöliittymä, johon on integroitu useita toisiaan tukevia modaliteetteja. Modaliteetilla viitataan tässä yhteydessä Dannenbergin ja Blattnerin (1992, xxiii) mukaisesti siihen aistiin, jolla informaatio havaitaan. Brownin, Newsomen ja Glinertin (1989, 339) mukaan visuaalisen informaation liiallinen määrä voi laskea tehokkuutta tehtävän suorituksessa. Jakamalla informaatiota useille eri aisteille voidaan näköaistin ylikuormitusta lievittää. (Brown ym. 1989, 339.) Huomioitavaa on silti se, että vaikka lähes kaikki käyttöliittymät ovatkin nykyään multimediakäyttöliittymiä, ne eivät silti ole multimodaalisia, koska hyödyntävät vain grafiikkaa (Brewster 1994, 8). Laitteisto on jo valmiina, kunhan vain tiedettäisiin mitä sillä pitäisi tehdä.

Tässä tutkimuksessa käyttöliittymää tarkastellaan yleisellä tasolla, ei laite- tai järjestelmätasolla (vrt. esimerkiksi Brewster, Leplatre & Crease (1998), jotka ovat tutkineet äänen käyttöä liikkuvan tietojenkäsittelyn sovelluksissa). Käyttäjän oletetaan kommunikoivan koneen kanssa perinteisesti näppäimistön ja hiiren (tai muun osoittimen) avulla (input) ja saavan palautetta äänen avulla (output), Brewsterin (1991, 18) tapaan. Näin aihetta voidaan tarkastella poikkitieteellisesti, pääpainon ollessa ihmisen ja tietokoneen vuorovaikutuksessa. Vuorovaikutuksen onnistuminen riippuu paljolti siitä, saako ihminen koneen tekemään halutun tehtävän pienimmällä mahdollisella käytön opettelemisella (Hereford & Winn 1994, 212-213). Tutkimuksessa ei käsitellä äänen käyttöä jaetuissa työtiloissa tai tietokoneavusteisessa ryhmätyössä (ks. Ackerman, Starr, Hindus & Mainwaring 1997).

Koska vuorovaikutuksen täytyy käyttöliittymässä olla nopeaa, äänten täytyy (a) olla lyhyitä ja (b) soida samanaikaisesti (toisin kuin peleissä tai elokuvissa, joissa immersiivisyys on tärkeämpää kuin tehokkuus; nämä eivät kuitenkaan välttämättä sulje toisiaan pois). Toisaalta Kramerin (1994, 4) mukaan ääni voi myös olla pitkä, jos sen tarkoituksena on välittää informaatiota taustaprosessista. Oli kyseessä sitten jääkaapin hurina, liikenne tai linnunlaulu, mukaudumme pysyviin, staattisiin äänielementteihin. Kuulemme yksittäiset äänet vasta keskityttyämme niihin tietoisesti tai kun ne vaativat huomiotamme yhtäkkisen muutoksen kautta. (Kramer 1994, 4.)

Ihmisellä on kyky erotella ja sijoittaa äänet tietyistä suunnista tuleviksi (Arons 1992, 35). Tämä ns. cocktail party effect (kyky valita yksi äänivirta useista samanaikaisista äänivirroista, Gibson 1966, 84) mahdollistaa äänten spatiaalisen sijoittelun käyttöliittymässä. Samalla tavalla kuin voimme näköaistin avulla havaita useita visuaalisia objekteja samanaikaisesti, voimme spatiaalisesti kuulla useita eri ääniobjekteja samanaikaisesti ja keskittyä johonkin niistä. *Spatialisaatiossa* luodaan synteettisesti kolmiulotteinen äänikenttä, jossa äänet

tuntuvat tulevan tietyistä paikallistettavista suunnista kuulijan pään ulkopuolelta (Goose & Möller 1999, 363). *Auralisaatio* taas tarkoittaa prosessia, jonka tarkoituksena on luoda jokin kolmiulotteinen virtuaalitila - joko sisä- tai ulkotilassa - simuloimalla äänilähteiden suuntaa ja huonekaikua (Begault 1994, 145). Käytännössä nämä kaksi käsitettä ovat hyvin lähellä toisiaan. Tässä tutkimuksessa auralisaatiolla tarkoitetaan koko virtuaalisen kuuloympäristön luontiprosessia, spatialisaation viitatessa enemminkin signaalinprosessointitekniikkaan.

Visuaalisia objekteja voidaan sijoittaa näytölle värien avulla, jolloin objektit erottuvat nopeammin toisistaan (Carter 1982, 127). Ääniobjektien (-tapahtumien) käytössä ja sijoittelussa sen sijaan on käytetty lähinnä ad hoc -menetelmiä riippuen käytettävästä laitteistosta, ohjelmistosta ja sovelluksen luonteesta. Ennen ääniobjektien sijoittelua tilassa täytyy käytettäville äänille luoda luokittelujärjestelmä, jonka tarkoituksena on antaa merkitys kullekin äänelle ja sen sijainnille tilassa.

1.2 Tutkimustehtävä

Aiemmissa tutkimustuloksissa esitetyt strukturoidut menetelmät äänen käytölle eivät ole olleet riittäviä, koska kukin näistä menetelmistä on keskittynyt vain johonkin kapeaan osa-alueeseen. TAULUKOSSA 1 on esitetty yleisen äänitutkimuksen kategoriat Schaferin (1977, 148) mukaan. Kendall (1995) lisää edelliseen vielä neuropsykologian, jossa tutkitaan kuulokokemuksen neurologisia rakenteita. Se, samoin kuin akustiikka ja äänen estetiikka kuuluvat tämän tutkimuksen ulkopuolelle (sikäli kuin näin tarkkoja rajauksia voi edes tehdä). Schafer (1977) on tehnyt akustisen äänimaiseman tutkimusta (soundscape), joka liittyy läheisesti myös akustiseen ekologiaan (acoustic ecology) (vrt. Gaver 1993a, 1993c).

TAULUKKO 1. Äänitutkimuksen kategoriat

ALUE	TUTKIJA	TUTKIMUSONGELMA
Äänen akustiikka	fyysikko, insinööri	Mitä äänet ovat?
Äänen psykoakustiikka	psykologi, fysiologi, kognitiotieteilijä	Miten äänet havaitaan?
Äänen semantiikka	kielitieteilijä, viestintätieteilijä	Mitä äänet tarkoittavat?
Äänen estetiikka	säveltäjä, musiikkitieteilijä	Miltä äänet tuntuvat?

Schafer (1977, 205) määrittelee akustisen ekologian tutkimusalaksi, jossa tarkastellaan äänimaiseman akustisten ilmiöiden suhdetta ja vaikutusta siinä elävien olentojen käyttäytymiseen. Määritelmän perusteella akustinen ekologia on täysin sovellettavissa käyttöliittymään, joka muodostaa oman vuorovaikutteisen äänimaisemansa.

Tämän tutkimuksen pääpaino sijoittuu pääasiassa psykoakustiikan, semiotiikan ja akustisen ekologian välimaastoon, jolloin saadaan toisaalta tietoa kuulokokemuksesta psyykkisenä prosessina, ja toisaalta tietoa ääniobjektien ominaisuuksista, sijainnista ja merkityksistä ympäröivässä maailmassa. Tämän jälkeen merkitykset voidaan liittää käyttöliittymäkontekstiin.

Keskeiseksi ongelmaksi muodostuu se, miten ääni voitaisiin integroida grafiikan kanssa. Äänen tulisi tukea tavoitteellista toimintaa, immersiivisyys voi käyttöliittymässä tulla kyseeseen vasta toiminnallisuuden varmistamisen jälkeen. Tavoitteena on löytää akustisesta ääniympäristöstämme hyödyllisiä spatiaalisia metaforia ja ikonisia ääniä piilossa olevan informaation organisoimiseen ja esittämiseen käyttöliittymässä yhdessä näköaistin välittämän informaation kanssa.

Kirjallisuuden pohjalta on tarkoitus koota yhteen äänitutkimukseen liittyviä käsitteitä ja teoreettisia viitekehyksiä. Tämän jälkeen käsitteitä tarkastellaan laajemmassa semioottisessa ja akustisen ekologian viitekehyksessä. Aihe on vaikea, koska yleistysten teko on hankalaa johtuen ihmisten kulttuuritaustoista sekä erilaisista äänen havainnointi- ja paikantamistaidoista. Aihetta tarkastellaan tässä tutkimuksessa normaalin näön omaavien keskivertokäyttäjien ja -havainnoijien näkökulmasta. Edellisen perusteella tutkimusongelma muodostuu seuraavasta kahdesta osaongelmasta:

1. Miten spatiaalista ääntä voitaisiin käyttää yhdessä grafiikan kanssa piilossa olevien objektien tai tapahtumien esittämisessä, tehtävän suorituksessa ja merkityksenannossa?
2. Voidaanko sellaista spatiaalista ääntä, joka ei selvästi viittaa käyttöliittymän tapahtumiin tai objekteihin, käyttää piiloinformaation välittämisessä?

Tutkimus jakaantuu siten kahteen pääteemaan: (a) spatiaalisen äänen tuottamaan merkityssisältöön ja (b) ympäristön spatiaalisten äänten hyödyntämiseen ja liittämiseen käyttöliittymän objekteihin. Taustaolettamuksena ovat kuuloaistin funktiot Gibsonin (1966, 75) mukaan: äänen suunnan havaitseminen ja äänilähteen tunnistaminen. Vaikka kyseessä on teoreettinen tutkimus, tutkimustulosten odotetaan hyödyttävän käyttöliittymäsuunnittelijoita myös käytännössä. Äänellinen käyttöliittymäsuunnittelu on aloitettava ihmisestä käsin,

ja edettävä vasta sitten tekniseen toteutukseen. Tutkimuksen hyödyllisimpänä puolena lienee se, että siinä pyritään kattamaan laajempi alue kuin aiemmissa tutkimuksissa - joskin verraten yleisellä tasolla - ja luomaan laajennettu viitekehys spatiaalisen äänen informaation välittämistä ja tavoitteellista toimintaa tukevasta käytöstä käyttöliittymässä.

1.3 Sovellettavat tieteenalat ja tutkimuksen rakenne

Tutkimuksessa käytetään kirjallisuutta soveltuvin osin akustisesta ekologiasta, kognitiivisesta psykologiasta ja viestintätieteistä. Viestintätieteiden osalta tutkimuksessa käytetään Fisken (2000) esittämää jakoa prosessikoulukuntaan ja semioottiseen koulukuntaan. Vaikka Fiske jakaa viestintätieteiden teoriat edellämainittuihin koulukuntiin, niiden välinen raja on usein häilyvä. Prosessikoulukunnan mukaan viestintä on sanomien siirtoa ja jos vaikutus eroaa tarkoitetusta, viestintä on epäonnistunut. Tällöin käydään läpi viestinnän eri vaiheita vian löytämiseksi.

Semioottisen koulukunnan mukaan viestintä taas on merkitysten tuottamista ja vaihtoa. Tärkeää on se, kuinka sanomat (useimmiten tekstit, tässä yhteydessä äänet) ja ihmiset toimivat vuorovaikutuksessa tuottaakseen merkityksiä. Väärinkäsitykset eivät välttämättä todista viestinnän epäonnistuneen, vaan ne saattavat johtua lähettäjän ja vastaanottajan kulttuurieroista. (Fiske 2000, 14-15.) Pääpaino tässä tutkimuksessa on semioottisessa lähestymistavassa. Kognitiotieteiden osalta keskitytään muistin asettamiin rajoituksiin sekä seriaaliin ja paralleeliin tiedonkäsittelyyn.

Tutkimus etenee seuraavasti: Luvussa 2 perehdytään siihen, miten kuuloaistin avulla saadaan tietoa maailmasta sekä verrataan kuulo- ja näköaistia keskenään. Luvussa 3 tarkastellaan niitä tekijöitä, joiden perusteella ääni paikannetaan tietystä suunnasta tulevaksi sekä esitetään synteettisen spatialisoinnin periaate ja

toteutukseen liittyvät ongelmat. Luvussa 4 analysoidaan spatiaalista äänimaisemaa käyttöliittymäkontekstissa. Luvussa 5 kootaan yhteen tutkimuksen keskeiset tulokset ja verrataan niitä aiempiin tutkimuksiin. Lopuksi yhteenvedossa pohditaan jatkotutkimusaiheita.

2 KUUNTELEMALLA TIETOA MAAILMASTA

Tässä luvussa tarkastellaan, miten saamme kuuntelemalla tietoa maailmasta yhdistämällä tätä tietoa aiempiin kokemuksiimme. Luku on synteesi kuuntelemisen psykoakustisista, havainnollisista, syntaktisista ja semanttisista tekijöistä, ja samalla perustelu semioottiselle lähestymistavalle.

2.1 Ääni ja sen eteneminen korvaan

Ääni on sekä semanttinen että fysikaalinen käsite (Beaudouin-Lafon & Gaver 1994, 49-50). Äänen akustisia piirteitä kuvataan fysikaalisilla suureilla (kuten taajuudella), kun taas elämyksellinen äänimaailma muodostuu lukuisista äänielämyksiin liittyvistä piirteistä (Jauhiainen 1995, 15). Jauhiaisen (1995, 26-27) mukaan *ääni fysikaalisena käsitteenä* on äänilähteestä pallomaisesti kaikkiin suuntiin etenevää molekyylien värähtelyä, joka etenee väliaineessa aaltomaisina tihentymä- ja harventumavaiheina. Eri aineiden rajapintoja kohdatessaan ääni "osittain heijastuu, osittain imeytyy toiseen aineeseen, absorboituu ja jatkaa kulkuaan toisessa aineessa" (Jauhiainen 1995, 37).

Äänen aaltoliikettä kuvataan aallonpituudella, taajuudella ja amplitudilla. Jauhiaisen (1995, 26-27) mukaan *aallonpituus* on etäisyys ääniaallon tihentymisvaiheesta toiseen, ja *taajuus* (frequency, yksikkönä hertsi, Hz) ilmoittaa värähdysten lukumäärän sekunnissa kuvaten äänen korkeutta. *Amplitudi* tarkoittaa värähtelyn laajuutta eli poikkeamaa keskiarvosta ja ilmaisee siten äänen voimakkuutta. *Äänes* eli *puhdas ääni* tarkoittaa ääntä, joka sisältää vain yhden taajuuden. (Jauhiainen 1995, 26-27.) Gibsonin (1966, 87) mielestä tällainen puhdas ääni aiheuttaa vain merkityksettömän aistimuksen. Merkitykselliset äänet koostuvat hänen mukaansa paljon monimutkaisemmista variaatioista. Jauhiainen (1995, 26-27) huomauttaa, että käytännössä kaikki

ympäristössä kuulemamme äänet koostuvat useasta samanaikaisesta eri taajuisesta värähtelystä. Tällaista ääntä kutsutaan seosääneksi, jonka taajuussisällön kuvausta kutsutaan äänen *spektriksi*. Spektri sisältää siis äänen perustaajuuden ja sen harmoniset osaäänekset kullakin hetkellä. (Jauhiainen 1995, 26-27; Goldstein 1999, 312-313.) Ääni voi olla jokin luonnollinen akustisen ilmiön ääni tai koneellisesti aikaansaatu, synteettinen ääni (Huopaniemi 1999, 43-44). Ääni etenee kuulijan korvaan sekä suoraan että heijastuen seinistä, lattioista ja katosta tai huonekaluista (Jauhiainen 1995, 38). Ääniaallot saavuttavat ensin korvalehden, joka ohjaa aallot korvakäytävän kautta tärykalvolle ja edelleen syvemmälle kuuloelimiin (Jauhiainen 1995, 43-44). *Ääni elämyksellisenä käsitteenä* sen sijaan koostuu neljästä peruspiirteestä:

- äänen voimakkuudesta kvantitatiivisena ominaisuutena (loudness)
- äänen korkeudesta kvalitatiivisena ominaisuutena (pitch)
- äänen ajallisuus- ja paikallisuusominaisuuksista.

Äänielämys muodostaa yksittäisen olion elämysavaruudessa, ja äänielämykset muuttuvat nopeasti ajassa ja seuraavat toinen toistaan (Jauhiainen 1995, 15-16.) Havaitsemamme äänimaailma (auditory scene, auditory space) muodostuu useista samanaikaisista äänivirroista (auditory stream), jotka voimme erottaa toisistaan ja joita voimme kuunnella selektiivisesti (Jauhiainen 1995, 19). Tässä yhteydessä voimme rinnastaa äänielämyksen äänivirtaan (ks. tarkemmin kohta 2.2). Jauhiaisen (1995, 18-19) mukaan äänen ajallisuus ja paikallisuus ovat tässä ja nyt koettuja, ja siten tärkeämpiä piirteitä elämyksessä kuin voimakkuus ja korkeus. Äänen paikantamisen avulla voimme mieltää itsemme suhteessa ympäröivään fyysiseen äänimaailmaan. (Jauhiainen 1995, 18-19.)

Kun kaksi korkeudeltaan ja voimakkuudeltaan samantasoista ääntä kuulostaa erilaiselta, kyse on erosta äänen *värissä* (timbre). Äänen väri muodostuu äänen syttymisen (attack), sammumisen (decay) ja äänen harmonisen rakenteen

perusteella. (Goldstein 1999, 318, 356.) Oleellista ei kuitenkaan ole äänen merkityksetön jakaminen korkeuteen, voimakkuuteen tai kestoon, vaan se, miten äänilähteet erotellaan toisistaan merkityksellisten tapahtumien havaitsemiseksi ja luokittelemiseksi hierarkioihin (Gibson 1966, 89). Jauhiaisen (1995, 15) mielestä lyhytkestoiset äänielämykset sisältävät rajoitetusti merkityssisältöä, kun taas pitempikestoiset äänielämykset muodostavat mielekkäitä kokonaisuuksia sisältäen paitsi tietoa äänilähteiden ominaisuuksista, myös äänen välittämiä käsitteellisiä, tunnepitoisia, esteettisiä, arvostuksellisia tai tahtomista ilmaisevia merkityksiä. Jauhiaisen näkemys on äärimmäisen yksinkertaistettu: se ei määrittele, mikä on lyhyt ja mikä pitkä ääni, eikä liioin ota huomioon kontekstin merkitystä. Lyhytkin ääni voi olla tietyssä kontekstissa erittäin merkitsevä.

Mansur ym. (1985, 117) jakavat äänen seuraaviin parametreihin: korkeus, voimakkuus, spatiaalinen sijainti, kesto, sointiväri, äänen syttyminen sekä ajoitus. Niin vähän kuin nämä erottelut kertovatkin äänten monimutkaisesta rakenteesta, ne auttavat hahmottamaan tutkimusaluetta ja jakamaan sen pienempiin kokonaisuuksiin. Tässä tutkimuksessa keskitytään ainoastaan äänielämyksen ajallisuuden ja paikallisuuden dynaamiseen suhteeseen sekä niiden sisältämään informaatiosisältöön (merkitykseen) käyttöliittymässä. Selvyyden vuoksi jatkossa puhutaan pelkästään äänestä (tai ääniobjektista/ -tapahtumasta, ks. kohta 2.4).

2.2 Äänten ryhmittely

Ryhmittelemme ääniä erillisiksi *äänivirroiksi* (auditory stream) (Kramer 1994, 4). Bregmanin ja Campbellin (1971, 244) mukaan äänivirta muodostuu samanlaisista äänitapahtumista, jotka erotellaan muista samanaikaisista äänitapahtumista. Bregman ja Campbell (1971, 244) olettavat, että kuuntelija voi suunnata huomionsa vain yhteen äänivirtaan kerrallaan. Äänivirta mentaalisena kokemuksena vastaa visuaalisen objektin kokemista ja on siten analoginen

suhteessa hahmopsykologiaan (Gestalt) (Jones 1993, 98; Smith, Pickett & Williams 1994, 174; Williams 1994, 97). Äänivirran ja aistikanavan (ks. Broadbent 1958) ero Bregmanin ja Campbellin (1971, 249) mukaan on siinä, että äänivirta on kullakin hetkellä järjestelty kokonaisuus, eikä sitä voida määritellä yhtenä fyysisenä ominaisuutena. Äänivirran funktiona on järjestellä äänimateriaali ensin, jonka jälkeen aistikanava voi prosessoida tätä äänimateriaalia yksi äänivirta kerrallaan (Bregman & Campbell 1971, 249). Chion (1990, 13), Ballas (1994, 80) ja Williams (1994, 96) painottavat, että äänivirta on havaintoon perustuva tulkinta äänilähteestä, eikä siis vastaa äänilähteen aiheuttamaa fysikaalista tapahtumaa. Williamsin (1994, 96-97) mielestä ideaalitilanteessa ollaan silloin, kun havaitsemamme äänivirta sisältää vain äänilähteen tunnistamisessa tarvittavan keskeisen informaation. Tämä on tärkeä huomio: turha informaatio voidaan poistaa suunniteltaessa ääniä käyttöliittymään.

Williams (1994, 96) määrittelee äänten ryhmittelyn havainnolliseksi prosessiksi, jossa kuuntelija erottelee akustisesta signaalista saadun informaation yksittäisiksi merkityksellisiksi äänitapahtumiksi. Kuuntelija voi kuunnella yhtä äänilähdettä ja samalla kuulla, mutta olla kiinnittämättä huomiotaan muihin äänilähteisiin (Gibson 1966, 83). Goldstein (1999, 359-362) esittää yleiset periaatteet äänten ryhmittelylle. Äänet ryhmitellään ensinnäkin niiden sijainnin perusteella. Äänet, jotka ovat peräisin yksittäisestä äänilähteestä, tulevat yleensä yhdestä sijainnista äänellisestä avaruudesta. Näin ollen voimme erottaa vasemmalla puolellamme käydyn hiljaisen keskustelun takanamme käydystä äänekkäästä keskustelusta osittain siksi, että ne sijoittuvat eri paikkoihin. Juuri paikantamisen avulla kykenemme havaitsemaan ja erottelemaan äänet toisistaan (luvussa 3 käsitellään tätä tarkemmin).

Äänen sijainti ei kuitenkaan ole keskeinen (eikä ainoa) tekijä erottelun kannalta: voimme kuunnella kamariorkesteria yhdestä kaiuttimesta ja silti selkeästi erotella

yksittäiset soittimet (Fitch & Kramer 1994, 308). Kuulohavaintojärjestelmämme pystyy siis erottelemaan useita äänivirtoja "akustisesta suosta" yhtä aikaa ilman että meidän täytyisi turvautua spatiaalisiin vihjeisiin (Fitch & Kramer 1994, 308). Goldsteinin (1999, 359-362) mukaan äänet ryhmitellään sijainnin lisäksi niiden äänenvärin, sävelkorkeuden ja esiintymistiheyden perusteella. Äänet jotka alkavat ja loppuvat eri aikaan, ryhmitellään eri ryhmiin, kun taas äänet, jotka ovat staattisia ja jatkuvia, ryhmitellään samasta äänilähteestä tulevaksi. (Goldstein 1999, 359-362.)

2.3 Aistihavaintojärjestelmät

Gibson (1966, 47) korostaa aistien olevan aktiivisia mieluummin kuin passiivisia, järjestelmiä enemmän kuin kanavia ja enemmän toistensa kanssa vuorovaikutuksessa olevia kuin toisensa poissulkevia. Gibson kutsuukin aisteja *aktiivisiksi havaintojärjestelmiksi* (perceptual systems). (Gibson 1966, 47.) Tämä on tärkeä huomio: on tehtävä ero kuulemisen ja kuuntelemisen välillä (samoin kuin näkemisen ja katsomisen välillä). Ihmisen kuulohavaintojärjestelmä mahdollistaa kuulemisen (passiivista), kun taas kuunteleminen on ihmisen aktiivista toimintaa (Gibson 1966, 49-51). Gibsonin (1966, 83) mielestä kuulohavaintojärjestelmä on kuuntelemista varten; kuuleminen sen sijaan on sattumanvaraista.

Kuulohavaintojärjestelmän funktiona ei siis ole vain kuulemisen mahdollistaminen, vaan aktiivinen äänilähteen suunnan havaitseminen (ks. luku 3), jolloin voimme suuntautua ääntä kohti (tai siitä poispäin!), sekä äänilähteen luonteen havaitseminen, jolloin kykenemme tunnistamaan sen (Gibson 1966, 75) (ks. luku 4).

2.4 Audiovisuaalinen sopimus

Chion (1990, 9) käyttää abstraktia käsitettä *audiovisuaalinen sopimus* (audiovisual contract) kuvaamaan äänellisen ja visuaalisen havainnon vuorovaikutteista suhdetta. Kuuntelija/katselija sulauttaa mielessään äänen ja kuvan yhdeksi kokonaisuudeksi (Chion 1990, 9). Kysymykset "mitä kuulen siinä minkä näen?" ja "mitä näen siinä minkä kuulen?" ilmentävät audiovisuaalista sopimusta (Chion 1990, 192). Seuraavassa tarkastellaan tätä kuulo- ja näköaistin suhdetta sekä kokonaisuutena että kummankin aistin kohdalla erikseen, jotta saadaan selville kunkin vahvimmat puolet. Tässä tutkimuksessa audiovisuaalinen sopimus viittaa äänen ja kuvan yhtäaikaiseen esittämiseen multimodaalissa käyttöliittymässä.

Kun edellisessä kohdassa määrittelimme kuuntelemisen aktiiviseksi toiminnaksi, on tarpeen pohtia sitä, miten se eroaa katselemisesta. Ensinnäkin, voimme kuunnella ja paikantaa sellaisia objekteja, joita emme näe (Kramer 1994, 4). Schaeffer (1967, 91-99) kutsuu tätä akusmaattiseksi kuuntelemiseksi (acousmatic listening) (Chion 1990, 71). Siinä missä näköaistin avulla keskitymme vain yhteen suuntaan ja saamme yksityiskohtaista tietoa silloisessa (rajoittuneessa) näköpiirissämme olevista objekteista, voimme kuuloaistin avulla tarkkailla ympäristöämme samanaikaisesti kaikista suunnista (Kramer 1994, 4). Tässä on ratkaiseva ero: enemmän kuin puolet senhetkisestä ympäristöstä jää näköpiirimme ulkopuolelle (Perrott, Saberi, Brown & Strybel 1990, 214). Perrottin ym. (1990, 214) mukaan tässä kapeassa näköpiirissä informaation prosessointikapasiteetti on verraten rajattu, ollen tehokkainta vain katseen keskipisteen ympärillä. Kuuloaistimme avulla voimme siis määritellä, mitkä osat ympäristöstämme tuodaan näköaistin alueelle (Gibson 1966, 83; Perrott ym. 1990, 214-215; Wenzel 1992, 81).

Perrott ym. (1990, 219) osoittivat tutkimuksessaan, että akustinen spatiaalinen informaatio auttaa löytämään näytöltä objektin nopeammin, kun kohde on näkökentän ulkopuolella. Vielä merkittävämpää tutkimuksessa on kuitenkin se, että myös näkökentän keskellä oleva kohde löytyi spatiaalisen äänen avulla nopeammin kuin pelkästään näön avulla (Perrott ym. 1990, 224). Heidän hypoteesinsa, että kuulohavaintojärjestelmän ensisijainen funktio on osoittaa silmille minne katsoa (Perrott ym. 1990, 214), saa siten vahvistusta. Äänen avulla voidaan siis saada tietoa taka-alan prosesseista. Perrott ym. (1990, 215) toteavat, että ihmiset painottavat kuitenkin enemmän visuaalista informaatiota kuin minkään muun aistin tuottamaa informaatiota. Tämä saattaa olla tunnusomaista länsimaiselle kulttuurille (jonka piirissä tietokoneiden kehitys jostain syystä tapahtuukin). Luultavasti tästä syystä käyttöliittymissä on käytetty yksinomaan visuaalista modaliteettia.

Toinen oleellinen ero suuntautumisen lisäksi on ajallisuus ja spatiaalisuus. Chionin (1990, 11, 136) sekä Fitchin ja Kramerin (1994, 308) mukaan näköaistin vahvuutena on objektien spatiaalinen havaitseminen, jolloin kuuloaistille jää niiden ajallinen ulottuvuus. Perrott ym. (1990, 219) sekä Fitch ja Kramer (1994, 322-323) olettavat, että näköaisti on kehityksen kuluessa tottunut prosessoimaan spatiaalisia objekteja peräkkäisesti (jolloin katselijan täytyy silmien liikkeellä käydä kaikki objektit läpi), kun taas kuuloaisti on erikoistunut useiden väliaikaisten objektien havaitsemiseen samanaikaisesti. Näkemys on vahvasti yksinkertaistettu: spatiaaliset ja ajalliset ulottuvuudet menevät aistien kesken myös limittäin, kuten saamme myöhemmin huomata (ks. myös Gibson 1966). Tällainen keinotekoinen erottelu on kuitenkin hyödyllinen, koska sen avulla päästään tarkastelemaan kuulon ja näön keskeisiä eroavaisuuksia. Erottelulla on myös fysikaaliset perusteet: Gibsonin (1966, 49-51) ja Gaverin (1989, 71) mukaan näkö perustuu valon heijastukseen erilaisista pinnoista, kun taas ääni on seurausta materiaalien värähtelystä (aiheuttaen ilman värähtelyä).

Emme voi kuulla ilmanpaineen absoluuttista arvoa, vain sen muutoksen ajassa (Scaletti & Craig 1991, 210). Äänen ajallisen ominaisuuden vuoksi se voidaan hyvin liittää animoituun grafiikkaan; staattinen visuaalinen ikoni ei välttämättä kuvaa ääntä parhaalla mahdollisella tavalla (Scaletti & Craig 1991, 210; Blattner ym. 1994, 455). Gaver (1989, 70-71) on päässyt asian ytimeen seuraavalla toteamuksellaan: "sound exists in time and over space, vision exists in space and over time." Hänen mukaansa äänet ovat olemassa ajassa, jolloin niiden alku ja loppu huomataan usein; sitä vastoin visuaaliset objektit ovat pysyviä ja niiden ilmestyminen tai häipyminen huomataan vain satunnaisesti. Ääntä ei välttämättä kuulla kuin kerran, mutta visuaalista objektia voi aina katsoa uudestaan. Tästä syystä ääni sopii erityisen hyvin kuvaamaan väliaikaisia tapahtumia (kausaaliset tapahtumat, esimerkiksi oven sulkeminen; ks. Fitch & Kramer 1994, 324), kun taas staattisia objekteja voidaan paremmin kuvata visuaalisesti. Toisaalta visuaaliset objektit saattavat tukkia näytön kokonaan. Gaver pohtii, että visuaalisia objekteja voidaan sijoittaa samanaikaisesti useisiin paikkoihin ja että spatiaalisten samanaikaisten äänten määrä on rajoitetumpi kuin visuaalisten objektien. (Gaver 1989, 71.)

Niin tärkeitä kuin Gaverin huomiot ovatkin, hänen näkökulmansa tässä suhteessa on selvästi rajoittunut ja aikansa tuote: tässä tutkimuksessa tullaan osoittamaan, että spatiaalisen äänen avulla voidaan onnistuneesti esittää useita objekteja samanaikaisesti, jolloin myös visuaalisten objektien suma näytöllä hälvenee. Ääni todellakin on kiinni ajassa; tästä on osoituksena se, ettei ääniä ole säilynyt menneiltä ajoilta, toisin kuin visuaalisia esineitä. Vasta äänen tallennustekniikoiden ansiosta ääntä on voitu vangita ja toistaa aina uudelleen.

Edellä on ollut puhetta visuaalisesta ja äänellisestä *objektista* ja *äänitapahtumasta*. Mikä sitten on näiden ero, ja miten objekti määritellään? Visuaaliset objektit voidaan nähdä ja niiden rajat ovat selvät. Fitchin ja Kramerin (1994, 308) mukaan objektin spatiaalinen sijainti ja osien yhtenäisyys

määrittävät visuaalisen objektin, jolloin näyttö voidaan helposti jakaa erillisiksi komponenteiksi. Ääniobjektin kohdalla tilanne on pulmallisempi: miten määritellä ääniobjektin rajat? Tähän on useita tapoja: (a) sijoittamalla äänet spatiaalisesti, (b) määrittelemällä objekti vuorovaikutuksessa toisen objektin kanssa, tai (c) Kramerin (1994, 12) tapaan joko liikuttamalla ääntä sijainnista toiseen tai liikuttamalla subjektia äänen läpi. Kramer (ks. myös Blattner ym. 1994, 454) siis ajattelee ääniobjektin rajaavan "palan" äänellisestä tilasta juuri liikkeensä avulla. Tässä tutkimuksessa ääniobjektien erottamiseksi on valittu spatiaalinen ja paikallaan pysyvä ääni.

Eräs pohdittava asia on myös se, onko ääniobjekti yksi ääni vai sarja nopeita, peräkkäisiä ääniä. Jones (1993, 98) kutsuu äänellistä tapahtumaa objektiksi, perustellen sen olevan analoginen suhteessa visuaaliseen objektiin. Schafer (1977, 129-130) puolestaan nimittää laboratorioissa toistettavia ääniä (siis kontekstistaan irrotettuja ääniä) ääniobjekteiksi, kun taas äänitapahtumiksi sellaisia ääniä, jotka tapahtuvat tietyssä paikassa tiettynä aikana ja aiheuttavat siten merkityksen. Tällöin ne viittaavat kontekstiin, jossa ne tapahtuvat, sisältäen alun, keskikohdan ja lopun. Teollisen ajan keinotekoiset äänet sisältävät lähes yksinomaan äänen keskikohdan ilman selkeää alkua tai loppua, toisin kuin luonnolliset äänet (Schafer 1977, 130).

Tässä tutkimuksessa käytetään tästä eteenpäin käsitteitä äänitapahtuma ja ääniobjekti rinnakkain, jälkimmäisen viitatessa sen visuaaliseen vastineeseen. Oleellista on se, mitä ääni tarkoittaa eri konteksteissa ja se, että peräkkäisten äänten järjestys on looginen. Tähän asiaan palataan kohdassa 2.6. Äänimaisema muodostuu vuorovaikutteisista äänitapahtumista, jotka puolestaan syntyvät erilaisten vuorovaikutusten tuloksena (Blattner ym. 1994, 454). Blattnerin ym. (1994, 454) mielestä assosioimalla spatiaalinen ääni vastaavaan visuaaliseen objektiin myös ääni muistetaan paremmin.

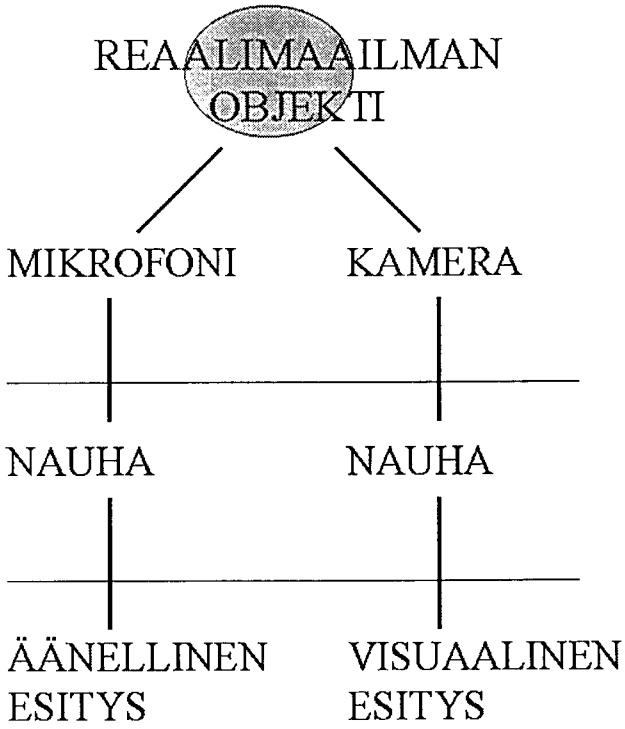
Ääni siis perustuu vaihteluihin ajassa ja viittaa yleensä muutokseen ja tilan tarkkailuun (Bly 1982, 374; Wenzel 1992, 81; Kramer 1994, 4; Walker & Brewster 2000, 2), kuvan ollessa joko staattinen tai liikkuva (animoitu). Esimerkkinä tilan muutoksesta on viinipullosta lähtevä ääni sen tyhjentyessä. Chionin (1990, 9-10) mielestä kuitenkin myös ääni voi joissain tilanteissa ilmaista staattisuutta; tällöin äänessä ei ole mitään variaatioita. Tällaisia ääniä ovat esimerkiksi puhelimen valintaääni, kaiuttimen hurina tai jokin äänisilmukka. On vaikea olla kuulematta äänessä edes vähäisiä vihjeitä epäsäännöllisyydestä tai liikkeestä. (Chion 1990, 9-10.) Käyttöliittymäkontekstissa on kuitenkin tärkeää, ettei ääni ole staattinen, vaan että sillä on aina jokin selkeä funktio.

Edellä esitetyn perusteella päädytään esittämään Scalettin ja Craigin (1991, 210) käyttämä malli, johon myös tämä tutkimus pohjimmiltaan perustuu. Mallin ideana on, että reaalimaailman objektit voidaan esittää monella tavalla, tässä tapauksessa joko äänen tai kuvan avulla. Objekteilla voidaan olettaa olevan sekä visuaalisia että äänellisiä ulottuvuuksia (Blattner ym. 1994, 465). Olettamuksena siis on, että katseleminen täydentää kuuntelemista ja toisinpäin: tästä seuraa, että saman objektin tarjoama äänellinen tai kuvallinen informaatio voi tilanteesta riippuen olla myös toisteista. KUVIOSSA 1 on havainnollistettu tätä mallia. Seuraavassa kohdassa tarkastellaan kahta erilaista kuuntelemisen tapaa.

2.5 Jokapäiväinen ja musiikillinen kuuntelu

Gaver (1989, 72-73; 1993a, 286; 1993c, 1-2) tekee eron *musiikillisen* ja *jokapäiväisen kuuntelemisen* välillä: edellinen viittaa ääniaallon fyysisten attribuuttien havaitsemiseen (esimerkiksi äänen korkeus), ja jälkimmäinen äänilähteiden havaitsemiseen ympäristössämme. Gaverin mielestä emme oven sulkeutuessa kuule siitä aiheutuvan äänen korkeutta, vaan sen sijaan saatamme havaita oven koon, materiaalin ja sulkemiseen käytetyn voiman. (Gaver 1989, 72-73.) Keskeistä Gaverin ajattelussa on se, että jokapäiväisessä elämässämme

emme kuule ääniä itsessään, sellaisenaan, vaan äänilähteen ominaisuuksia tietyssä kontekstissa. Näkökulma perustuu Gibsonin (1966) ajatuksiin. Chion (1990, 25) kutsuu tätä kausaaliseksi kuuntelemiseksi. Chionin (1990, 26-27) mielestä kontekstistaan irrotetun yksittäisen äänen tunnistaminen on vaikeaa: saatamme tunnistaa vain äänilähteen yleiset piirteet, esimerkiksi "jonkin koneellisen äänen". Chionin mielestä voimme kuulla äänen kausaalisesti, eli siis kuulla ja tunnistaa raapimisen tai aistia sen nopeuden, mutta sen sijaan emme kuule mikä raapii mitä (Chion 1990, 26-27).



KUVIO 1. Objektin äänellinen ja visuaalinen esitys

Chion (1990, 27) painottaa, että ääni on seurausta vähintään kahdesta äänilähteestä. Chion siis erottaa äänilähteen vuorovaikutuksen muodosta ja eroaa siten hieman Gaverin teoriasta. Chion (1990, 28) käyttää vielä nimitystä

semanttinen kuunteleminen, joka tarkoittaa viestin tulkitsemista tietyn koodin tai kielen avulla. Käytännössä semanttinen ja jokapäiväinen kuunteleminen tarkoittaa samaa asiaa: kyse on äänten tulkitsemisesta. Gaver (1993c, 18) painottaa, ettei jokapäiväistä kuuntelua voi tutkia pelkästään äänilähteiden fyysisten ominaisuuksien perusteella; tärkeintä on se, mitä kuuntelijat havaitsevat.

Psykologit ovat perinteisesti olleet kiinnostuneita juuri musiikilliseen kuunteluun vaikuttavista havainnollisista ilmiöistä, vaikka suurin osa jokapäiväisestä kuuntelusta ei sitä ole (Gaver 1993c, 1-2). Gaver nimittää näitä perinteisen psykoakustisen tutkimuksen parametreja äänen primitiivisiksi (fysikaalisiksi) ulottuvuuksiksi (Gaver 1993c, 1-2). Schaeffer (1967, 270) kutsuu tätä pelkistetyksi kuuntelemiseksi (reduced listening), koska siinä keskitytään ääneen itsessään, irrotettuna sen syystä ja merkityksestä (Chion 1990, 29). Myös Chionin (1990, 31) mielestä äänellä on kausaalisuutensa lisäksi myös oma esteettinen arvonsa johtuen äänen väristä ja tekstuurista. Gaver (1993c, 1-2) painottaa kuitenkin, että ero on kokemuksessa, ei äänissä. On siis hyvinkin mahdollista kuunnella myös jokapäiväistä äänimaisemaamme musiikkina. Gaver mainitsee tässä säveltäjä John Cagen, joka on hyödyntänyt tätä seikkaa sävellyksissään. (Gaver 1993c, 1-2.)

Gaverin mielestä jokapäiväisen kuuntelun tuottama informaatio voi toimia perustana uuden viitekehyksen luomisessa: voimme manipuloida ääntä sen äänilähteen ulottuvuuksien kautta, ei itse äänen ulottuvuuksien kautta (Gaver 1993c, 2). Gaver vertailee ääniä seuraavalla tavalla: musiikilliset äänet eivät edusta päivittäin kuulemaamme äänimaisemaa, koska useimmat musiikilliset äänet ovat harmonisia, kun taas jokapäiväiset äänet eivät ole, sisältäen myös hälinää; musiikilliset äänet paljastavat vähemmän lähteistään kuin jokapäiväiset äänet; musiikillisilla instrumenteilla muutokset äänen korkeudessa tai voimakkuudessa (siis fyysisissä attribuuteissa) eivät ole niin informatiivisia kuin

jokapäisten äänten tuottamat variaatiot. (Gaver 1993c, 3.) Tämä on mielenkiintoinen näkökulma, koska tällöin päästään tarkastelemaan aktiivisesti äänimaisemaa ajassa, eikä pelkästään passiivisesti "kivettynyttä" musiikillista ärsykettä. Gaverin erittely paljastaa, että perinteiset musiikilliset äänet edustavat vain muutamaa prosentin murto-osaa siitä mahdollisesta kapasiteetista, jonka jokapäiväinen ääniympäristömme joka hetki tuottaa.

2.6 Syntaktiset ja semanttiset tekijät äänten tunnistamisessa

Käyttöliittymässä on tärkeää, että äänen merkitys on yksiselitteinen; tätä varten äänen on oltava tunnistettava. Kieli voidaan jakaa kolmeen tasoon: sanoihin (lexical level), kielioppiin/syntaksiin (syntactic level) ja semantiikkaan (Blattner ym. 1994, 450). Jos näitä tasoja sovelletaan ei-puheääniin, äänen fysikaaliset attribuutit (esimerkiksi spatiaalisuus) edustavat alinta tasoa - jolle suurin osa äänitutkimuksesta on keskittynyt (Blattner ym. 1994, 450). Ääniä voidaan prosessoida ylhäältä alas (top-down) tai alhaalta ylös (bottom-up). Howardin ja Ballaksen (1980, 431) mukaan ylhäältä alas prosessoiminen perustuu tietoon (knowledge-driven), kun taas alhaalta ylös dataan (data-driven). Esimerkiksi puheen havaitsemisessa kuuntelijat käyttävät sekä kielen syntaktista ja semanttista rakennetta että äänilähteestä tulevaa havainnollista informaatiota. Vähemmän on kuitenkin tutkittu syntaktisten ja semanttisten tekijöiden vaikutusta monimutkaisiin ei-puheääniin. (Howard & Ballas 1980, 431.)

Alhaalta ylös prosessoitaessa havaitaan äänen akustisia ominaisuuksia, joista rakennetaan merkityksiä muistin avulla, kun taas ylhäältä alas prosessoitaessa tulkitaan äänen merkitys sen kontekstin ja aiempien kokemusten perusteella, josta edetään äänen yksittäisten parametrien analysointiin (Ballas & Howard 1987, 96; Kramer 1994, 5; Goldstein 1999, 376; Anderson 2000, 63).

Monilla jokapäiväisillä monimutkaisilla ei-puheäänillä on määriteltävissä oleva jaksollinen rakenne (siis syntaksi kielen tavoin), samoin kuin semanttinen sisältö, ja ne havaitaan samanlaisten prosessien kautta kuin puhe (Howard & Ballas 1980, 431; Mansur ym. 1985, 117; Ballas & Howard 1987, 96; Blattner ym. 1994, 448). Erojakin toki on: kieli viittaa sosiaalisiin merkityksiin, kun taas ympäristön ei-puheäänet viittaavat kausaalisiin tapahtumiin (sidottu akustiikan fysikaalisiin lakeihin) (Gibson 1966, 91; Ballas & Howard 1987, 98). Esimerkiksi oven avaaminen ja henkilön astuminen huoneeseen muodostavat jakson ajallisesti järjestettyjä äänitapahtumia. Tällaisilla jaksoilla on tilapäinen rakenne, koska ne koostuvat yksittäisistä äänitapahtumista, jotka syntyvät äänilähteiden sanelemassa järjestyksessä ja joilla on äänilähteistä riippuva kesto. (Howard & Ballas 1980, 431.)

Myös Deutschin (1980, 381) tutkimus tukee tätä näkemystä rakenteen merkityksestä: voimme prosessoida monimutkaistakin informaatiota sarjallisesti, jos informaatio on systemaattisesti organisoitu ja havaitsija ymmärtää sen rakenteen. Ryhmittelemme sarjallisia kuvioita jaksoihin, joita sitten järjestelemme hierarkioihin. Esimerkiksi kielessä merkityksellisen lauseen muodostavat sanat havaitaan ja muistetaan paremmin kuin vain sanat sattumanvaraisessa järjestyksessä. (Deutsch 1980, 381.) Tämä seikka on keskeinen sijoitettaessa ääniä käyttöliittymän toimintoihin.

Ei-puheäänten kohdalla syntaksi syntyy niistä mahdollisista väliaikaisista suhteista, joista sarja äänitapahtumia voi keskenään muodostaa järkevän kokonaisuuden (esimerkiksi a-a-c-d-d) (Howard & Ballas 1980, 436). Howardin ja Ballaksen (1980, 436, 438) mukaan tämän kokonaisuuden on oltava sekä syntaktisesti että semanttisesti järkevä: syntaksin täytyy pohjautua äänitapahtumien väliaikaiseen rakenteeseen, ja yksittäisten äänitapahtumien täytyy noudattaa syntaksia. Muutoin kuuntelijoilla on vaikeuksia havaita rakennetta. (Howard & Ballas 1980, 436, 438.) Tätä seikkaa voidaan hyödyntää

jakamalla käyttöliittymän toiminnot äänitapahtumien jaksoiksi: jakso koostuu toiminnoista, jotka käyttäjä tekee suorittaakseen jonkin tehtävän.

Warrenin ja Verbruggen (1984) tavoitteena oli löytää ne tekijät, joiden perusteella kuuntelija tunnistaa eri vuorovaikutuksen muodon. He tutkivat kahta eri tapausta, pullon särkymistä ja pomppimista kovalla alustalla, ja huomasivat, että spektraaliset erot eivät olleet ratkaisevia tunnistamisessa. Kuuntelijat kykenevät erottamaan pullon pomppimisen ja särkymisen toisistaan vain niiden ajallisten jaksojen perusteella. (Warren & Verbrugge 1984, 711.) Tämä tutkimus on keskeinen, koska se vahvistaa edellä käsiteltyä näkemystä äänitapahtumien väliaikaisen rakenteen merkityksestä. Pomppiva pullo aiheuttaa erilaisen väliaikaisen jakson äänitapahtumia (tässä tapauksessa iskuja) kuin särkyvä pullo. Kuuntelijat reagoivat rytmiin halukkaammin kuin mihinkään muuhun äänen parametriin (Blattner, Sumikawa & Greenberg 1989, 23). Luultavasti tästä syystä väliaikaisen rakenteen merkitys on niin ratkaiseva tunnistamisessa.

Voimme käyttää myös kontekstia apuna objektien tunnistamisessa ja merkityksenannossa (Ballas & Howard 1987, 98; Ballas & Mullins 1991; Anderson 2000, 63). Usein hyvinkin erilaiset fysikaaliset tapahtumat aiheuttavat samankaltaisia ääniä (Blattner ym. 1994, 450). Schafer (1977, 150) ottaa esimerkin: käärmeen sihinä ja kattilan kiehuminen ovat äänenä lähellä toisiaan, mutta herättävät eri kontekstissa erilaisen kokemuksen. Kun nämä kaksi ääntä irroitetaan kontekstistaan ja toistetaan nauhalta, niiden identiteetti hämärtyy. Korva ei ole tarpeeksi tarkka erottamaan näiden kahden äänen fysikaalisia eroja erottaakseen käärmeen sihinän ja veden kiehumisen toisistaan nauhalta. (Schafer 1977, 150.) Toisena esimerkkinä voisi olla ääni ""klik-klik": tällainen ääni voi kuulua kuulakärkikynästä, valokatkaisijasta, nitojasta tai kamerasta. Ainoastaan konteksti voi poistaa epäilyksen siitä, mistä on kyse. Ballas (1993) arvioi akustisia, ekologisia, havainnollisia ja kognitiivisia tekijöitä, jotka ovat keskeisiä jokapäiväisten äänten tunnistamisessa. Tutkimuksessa oli mukana 41 lyhyttä ja

toisistaan eroavaa jokapäiväistä ääntä. Tutkimus osoitti, että äänilähteen tunnistamiseen vaikuttavat ainakin akustiset (fysikaaliset) muuttujat, ekologinen frekvenssi, kausaalinen epävarmuus ja äänen tyypillisyys (Ballas 1993, 262). Tutut äänet siis tunnistetaan nopeammin (Ballas & Howard 1987, 103; Ballas 1993, 256).

Edellä on alustavasti käsitelty äänten havaitsemiseen ja äänilähteiden tunnistamiseen liittyviä asioita. Luvun lopuksi on tarpeen koota keskeiset huomiot. Kuulohavaintojärjestelmän ansiosta voimme aktiivisesti suuntautua äänivirtaa kohti, kohdistaa katseemme tähän objektiin sekä tunnistaa sen joko kuulo- tai näköaistimme avulla. Jokapäiväiset äänet voivat välittää hienovaraisempaa informaatiota äänilähteestä kuin (perinteisten) musiikillisten instrumenttien tuottamat äänet. Kuuloaistilla tunnistettaessa on varmistettava, että äänitapahtumien sarja noudattaa syntaksia: tällöin äänitapahtumien muodostama kokonaisuus on semanttisesti looginen ja ymmärrettävä.

Anderson (2000, 81) summaa luvun vielä seuraavasti: valitsemme äänen, johon suuntaudumme, äänen fysikaalisten ominaisuuksien perusteella (esimerkiksi korkeuden perusteella), jolloin "voimistamme" kyseistä ääntä ja vaimennamme muut äänet. Nämä muut äänet eivät kuitenkaan kokonaan vaimennu, vaan kiinnittävät huomiomme jos ne ovat fyysisiltä ominaisuuksiltaan korostavia (esimerkiksi kova ääni), semanttiselta sisällöltään kiintoisia (jos nimemme mainitaan) tai yhteneväisiä sillä hetkellä prosessoimamme äänen kanssa. (Anderson 2000, 81.) Seuraavassa luvussa keskitytään tarkemmin yhteen äänen parametreista. Spatiaaliset jokapäiväiset äänet toimivat analogiana kielen sanoihin ja voivat muodostaa syntaksin - ja siten merkityksen - käyttöliittymäkontekstissa.

3 SPATIAALINEN ÄÄNI

Ääni välittää informaatiota äänilähteen ja ympäristön lisäksi myös sijainnista (Gaver 1993c, 7). Visuaalisen näkyvyyden heiketessä - esimerkiksi huonon valaistuksen vuoksi - täytyy turvautua spatiaaliseen äänimaisemaan (Goldstein 1999, 242). Samassa tilanteessa ollaan myös silloin, kun näyttö on liian täynnä visuaalista informaatiota. Kuten edellisessä luvussa todettiin, kuuntelija voi erottaa äänivirtoja myös yhdestä kaiuttimesta. Ääniä ryhmitellään siis myös muiden muuttujien kuin sijainnin perusteella. Onko spatiaalisen äänen funktio siten vain immersiivisyyden lisääminen tai navigointi tilassa? Tässä luvussa tarkastellaan spatiaalista ääntä ja pyritään löytämään tapoja sen hyödyntämiseen käyttöliittymässä. Jokapäiväisen spatiaalisen äänimaisemamme hyödyntäminen käyttöliittymässä voi parhaimmillaan johtaa intuitiiviseen vuorovaikutukseen.

Spatiaalisuus pitää sisällään useita tekstuureita: *polyfonia* muodostuu kahdesta tai useammasta samanaikaisesta, yhtä tärkeästä melodiasta; *homofonia* taas yhdestä päämelodiasta ja useista muista, säestävistä melodioista (Blattner, Greenberg & Kamegai 1992, 101). Käyttöliittymän kokonaisäänimaisema voisi siten muodostua polyfonisista äänivirroista sijoiteltuna spatiaalisesti. Äänimaisema voisi osin olla myös homofoninen, riippuen äänten tärkeydestä ja luokittelusta käyttöliittymän eri konteksteissa. Ennen spatiaalisen äänen tarkempaa määrittelyä on tarpeen käsitellä niitä tekijöitä, joiden perusteella ääni paikannetaan tietystä suunnasta tulevaksi. Luvussa pohditaan myös spatialisoinnin toteutukseen liittyviä ongelmia ja esitellään spatiaalista ääntä hyödyntäviä sovelluksia. Spatialisointia ja auralisaatiota voidaan tämän jälkeen tarkastella semioottisessa viitekehyksessä kuuntelijan näkökulmasta luvussa 4.

3.1 Äänen paikantaminen

Jauhiainen (1995, 150) määrittelee *lokalisaation* äänen paikantamiseksi ilman kuulokkeita ja *lateralisaation* äänen paikantamiseksi kuulokkeita käyttämällä. Lateralisaatio viittaa äänen paikantumiseen sivuun keskipisteestä, eli sijoittumista keskipisteestä jompaan kumpaan korvaan (Jauhiainen 1995, 150). Seuraavassa tarkastellaan niitä tekijöitä, joiden perusteella ääni paikannetaan tietystä suunnasta tulevaksi.

Vaihe-ero korvien välillä (interaural delay time, IDT) on merkittävä tekijä paikantamisessa (Burgess 1992b, 54). Vaihe tarkoittaa värähtelyn ajankohtaa tiettynä hetkenä. Jos viivettä ei ole, ääni tulee suoraan edestä, takaa tai yläpuolelta. Jos ääni tulee sivulta, ääniaallot tulevat perille eri aikaan eri korvaan. Viive voi enimmillään olla 0.63 millisekuntia, jonka aikana ääni kulkee pään läpimitan pituisen matkan. (Burgess 1992b, 54.) Äänen voimakkuusero korvien välillä (interaural intensity difference, IID) on toinen merkittävä tekijä äänen paikantamisessa (Kendall 1995).

Jos ääni tulee sivusta, sen täytyy ennen toiseen korvaan etenemistään kiertää ensin pään ympäri. Pää imee itseensä osan ääniaalloista, joten alkuperäinen ääni tulee toiseen korvaan hieman vaimennettuna. Tätä ilmiötä kutsutaan pään aiheuttamaksi äänen vaimenemiseksi (head shadow effect). (Burgess 1992b, 54.) Samoin äänen tullessa takaa korvalehdet vaimentavat ääntä hieman verrattuna siihen, että ääni tulisi edestä (Goose & Möller 1999, 365). Alle 1500 Hz:n taajuiset ääniaallot ovat jo niin pitkiä, että ne taipuvat pään ympäri, jolloin voimakkuuseroa on vaikea havaita; paikantaminen tapahtuu tällöin pääasiassa vaihe-eron perusteella. Yli 1500 Hz:n taajuudet taas heijastuvat poispäin ja vaimentuvat pään vaikutuksesta, jolloin paikantaminen tehdään pääasiassa voimakkuuseron mukaan. 1500 Hz:n raja johtuu siitä, että 1500 Hz:n ääniaallon

aallonpituus on sama kuin pään läpimitta. (Kendall 1995.) Havaitsemme äänen sijainnin sivusuunnassa siis sekä vaihe- että voimakkuuseron perusteella.

Korvalehden ja -käytävän merkitys (pinna and ear canal response) äänen paikantamisessa on kaksikorvaisuuden ohella ratkaiseva. Koska jo vaihe- ja voimakkuuseron perusteella paikannamme ääniä vaaka-akselilla, korvalehtiä tarvitaan oikeastaan vain äänten paikantamiseen pystyakselilla. Tämän lisäksi korvalehtiä tarvitaan luomaan vaikutelma siitä, että ääni todellakin tulee pään ulkopuolelta eikä sisältä. Korvalehdet ohjaavat äänen korvakäytävään, mutta samalla korvalehtien epäsäännöllinen muoto muuttaa äänen spektriä vaimentamalla ja voimistamalla tiettyjä taajuuksia äänen tulosuunnan ja kunkin ihmisen yksilöllisen anatomian mukaan. Tämän jälkeen aivot vertaavat ja tulkitsevat saatua tietoa äänen paikantamiseksi. (Kendall 1995.)

Ihmisen yläruumis ja olkapäät heijastavat tiettyjä taajuuksia (shoulder echoes). Nämä heijastukset saapuvat korviin hieman myöhemmin riippuen äänen tulokulmasta. Myös tämä informaatio auttaa äänen tulosuunnan havainnoimisessa. (Burgess 1992b, 54.) On luonnollista, että käännämme päätämme halutessaan tarkemmin selvittää äänen suunnan. Jos käännämme päätämme sivulta tulevaa ääntä kohden, koemme äänen nyt tulevan keskeltä. Pään liikkeen avulla on siis mahdollista paikantaa ääni tarkemmin. Wun, Duhin, Ouhyoungin ja Wun (1997, 17) tutkimuksen mukaan pään kääntäminen äänen suuntaan parantaa paikantamista 90 % verrattuna pään pitämiseen paikallaan. Esimerkiksi jos ääni tulee etuoikealta, vaihe- ja voimakkuuserot ovat samat, jolloin äänilähde voisi teoriassa olla "peilikuvaäänenä" myös takaoikealla (Begault 1991, 866). Tilanne aiheuttaa epätietoisuutta, jos pää pidetään paikallaan. Pään kääntäminen oikealle poistaa heti peilikuvaäänen vaihe- ja voimakkuuserojen muuttuessa. Tällöin häviää epätietoisuus siitä, tuleeko ääni edestä vai takaa.

Myös näköaisti vaikuttaa paikantamiseen, joskin yllättävällä tavalla. Hylkäämme kuuloaistin tuoman informaation, jos se on ristiriidassa näkemämme äänilähteen sijainnin kanssa (Burgess 1992b, 55; Goldstein 1999, 242). Uskomme siis mieluummin näkö- kuin kuuloaistia.

Huopaniemen (1999, 34) mukaan *huonekaiku* (room impulse response, RIR) voidaan jakaa suoraan ääneen, ensiheijastumiin ja jälkikaiuntaan. *Suora ääni* on ääni, joka saapuu korvaan suoraan heijastumatta ensin mistään pinnasta. *Ensiheijastumat* (early echo response) ovat ääniä, jotka saapuvat kuulijan korvaan 50-100 millisekunnin kuluessa äänen syntymisestä (Burgess 1992b, 55). Binauraalinen (kaksikanavainen) huonekaiku (binaural room impulse response, BRIR) riippuu kuulijan ja äänilähteen sijainnista sekä huoneen ominaisuuksista tietyllä hetkellä (Kleiner, Dalenbeck & Svensson 1993, 865).

Jälkikaiunta tarkoittaa ensiheijastuman jälkeisiä äänen heijastuksia. Huoneen kaikuisuutta mitataan jälkikaiunta-aikana (reverberation time), jota mitataan voimakkaalla lyhyellä äänipulssilla. Jauhiaisen (1995, 38) mukaan jälkikaiunta-aika on aika, joka kuluu äänitason pienenemiseen 60 dB:iin käytetyn äänipulssin huippuarvosta. Suora ääni tulee ensimmäisenä suoraan kuulijan korvaan, ja kaikki ensi- ja jälkiheijastumat saapuvat tämän jälkeen eri aikoihin ja eri suunnista. Suoran äänen perusteella saamme vihjeitä äänilähteen suunnasta ja heijastumien määrän sekä äänen voimakkuuden perusteella äänen etäisyydestä ja tilan ominaisuuksista. (Burgess 1992b, 57; Kendall 1995; McGrath 1995, 1-2.)

Äänilähteen liikkuessa sen suunta ja etäisyys muuttuvat. *Doppler-ilmiö* (doppler-effect) syntyy, kun kohteen lähestyessä ääniaallot tihentyvät, jolloin aallonpituus lyhenee ja äänen korkeus nousee. Vastaavasti kohteen loitontuessa ääniaallot harventuvat, jolloin äänen korkeus laskee. Tästä voimme päätellä, että äänilähde liikkuu. Mitä nopeammin äänilähde liikkuu suoraan meitä kohti tai meistä poispäin lähietäisyydellä, sitä suurempi on äänen korkeuden muutos. Näin

voimme saada tietoa äänilähteen etäisyydestä ja suunnasta. (Schafer 1977, 80; Begault 1994, 40.)

Äänen paikantamiseen vaikuttavat siis vaihe- ja voimakkuuserot korvien välillä, pään aiheuttama äänen vaimeneminen, korvalehden ja korvakäytävän muoto, heijastumat olkapäistä ja yläruumiista, pään liike, näköaisti, suora ääni, ensiheijastumat, jälkikaiunta-aika, äänilähteen liikkuminen ja Dopplerin ilmiö. Näiden tekijöiden vaikutukset äänen spektriin ovat mitattavissa siirtofunktion avulla. Siirtofunktion määrittelemisen jälkeen esitetään menetelmä, jolla ääneen lisätään keinotekoisesti vihjeitä tilasta ja suunnasta.

3.2 Siirtofunktio

Begaultin (1991, 864) mukaan *siirtofunktion* (head related transfer function, HRTF) avulla saadaan selville, miten äänen spektri muuttuu äänen kulkiessa pään ulkopuolisesta äänilähteestä ulkokorvan läpi tärykalvolle. Siirtofunktio on erilainen kummassakin korvassa, ja se määritetään mittaamalla tärykalvoon saapuvan äänen spektrin muutokset verrattuna äänilähteestä lähtevän äänen spektriin. Tällöin voidaan tietokoneen avulla laskea alkuperäisen signaalin ja koehenkilön korviin tulevan signaalin perusteella kyseisen henkilön siirtofunktiot kunkin äänilähteen suhteen. Siirtofunktio vaihtelee äänilähteen sijainnin ja kuulijan pään asennon välisen suhteen mukaan kullakin hetkellä. (Begault 1991, 864; Kendall 1995).

Aina ei ole mahdollista mitata jokaisen yksilöllisen käyttäjän siirtofunktioita. Myös yleisen, keskiarvoisen siirtofunktion avulla päästään todentuntuisiin tuloksiin paikantamisessa (parhaiten kuitenkin vaaka-akselilla) (Begault 1991, 865; Wenzel, Wightman & Kistler 1991, 358; Goose & Möller 1999, 365). Keskiarvoinen siirtofunktio on määritetty mittaamalla se suurelta määrältä ihmisiä ja laskemalla keskiarvo. Toinen mahdollisuus on käyttää keskimääräistä

paremman "paikantajan" siirtofunktiota (Begault 1991, 865). Mittaukset tehdään yleensä kaiuttomassa huoneessa eri suunnista tulevilla äänillä laajalla taajuusalueella (McGrath 1995, 3).

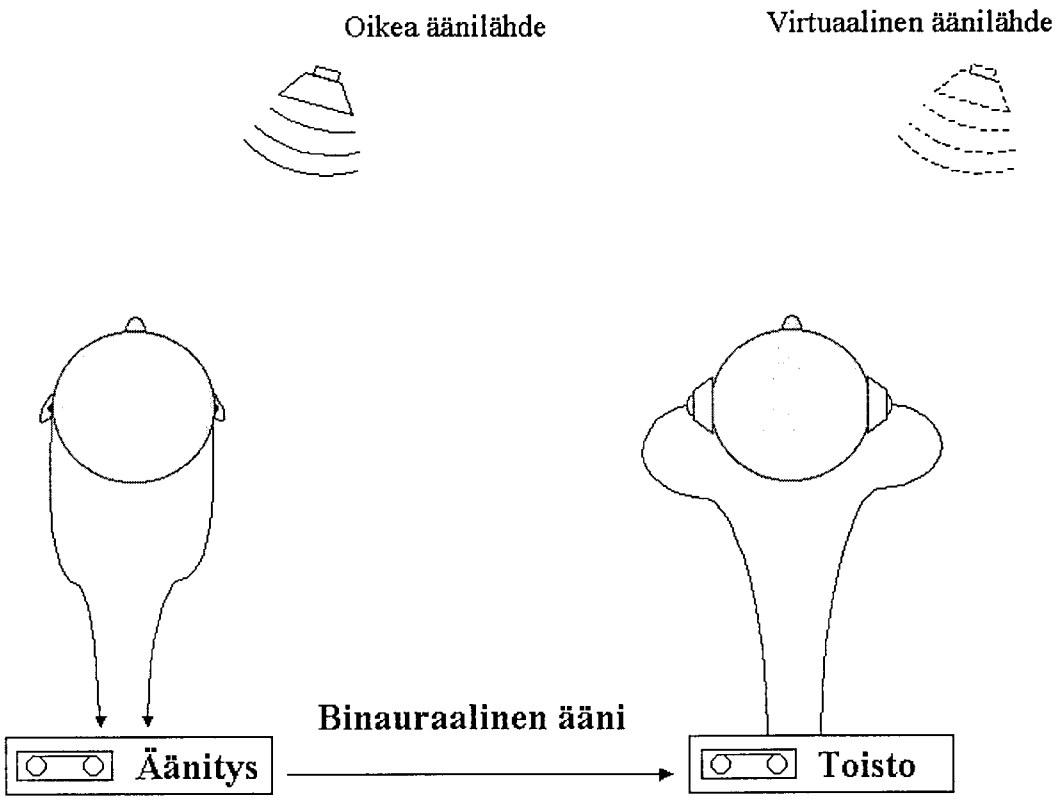
3.3 Äänen synteettinen spatialisointi

Spatialisointi tarkoittaa Goosen ja Möllerin (1999, 363) mukaan äänen prosessointia siten, että ääni koetaan tulevan kolmiulotteisesta äänikentästä (tietystä tunnistettavasta suunnasta) kuulijan ulkopuolelta. Burgessin (1992a, 2; 1992b, 53) mukaan digitaalinen monauraalinen ääni voidaan keinotekoisesti spatialisoida binauraaliseksi ääneksi digitaalisten suodatinalgoritmien avulla, kun tiedetään joko yksilöllinen tai keskiarvoinen siirtofunktio. Tätä kutsutaan synteettiseksi spatialisoinniksi.

Sekä stereoääni että binauraalinen ääni viittaavat kahden kanavan käyttöön, mutta poikkeavat äänitystekniikoiltaan ja äänentoisto-ominaisuuksiltaan toisistaan (monauraalinen viittaa yhden kanavan käyttöön). Burgessin (1992a, 1; 1992b, 53) mukaan stereoäänitys tarkoittaa monilla toisistaan erossa olevilla mikrofoneilla tehtyä äänitystä, jota kuunnellessaan kuulija voi havaita äänilähteen sijainnin äänikentässä. Tällöin sijainti on kuitenkin rajoitettu siihen vaakasuoraan linjaan, joka muodostuu kahden kaiuttimen välille tai kuulijan vasemman ja oikean korvan välille kuulokkeilla kuunneltaessa (ääni ei siis tunnu kuuluvan riittävästi pään ulkopuolelta) (Kendall 1995). Stereoäänitys ei kaksiulotteisena vastaa reaalitodellisuudessa korviimme tulevaa ääntä, koska se sisältää vain alkeellisen syvyysvaikutelman (lähellä - kaukana) ja leveysvaikutelman (vasen - oikea). Kolmiulotteinen ääni (three-dimensional sound) luo illuusion syvyyden ja leveyden lisäksi myös korkeudesta.

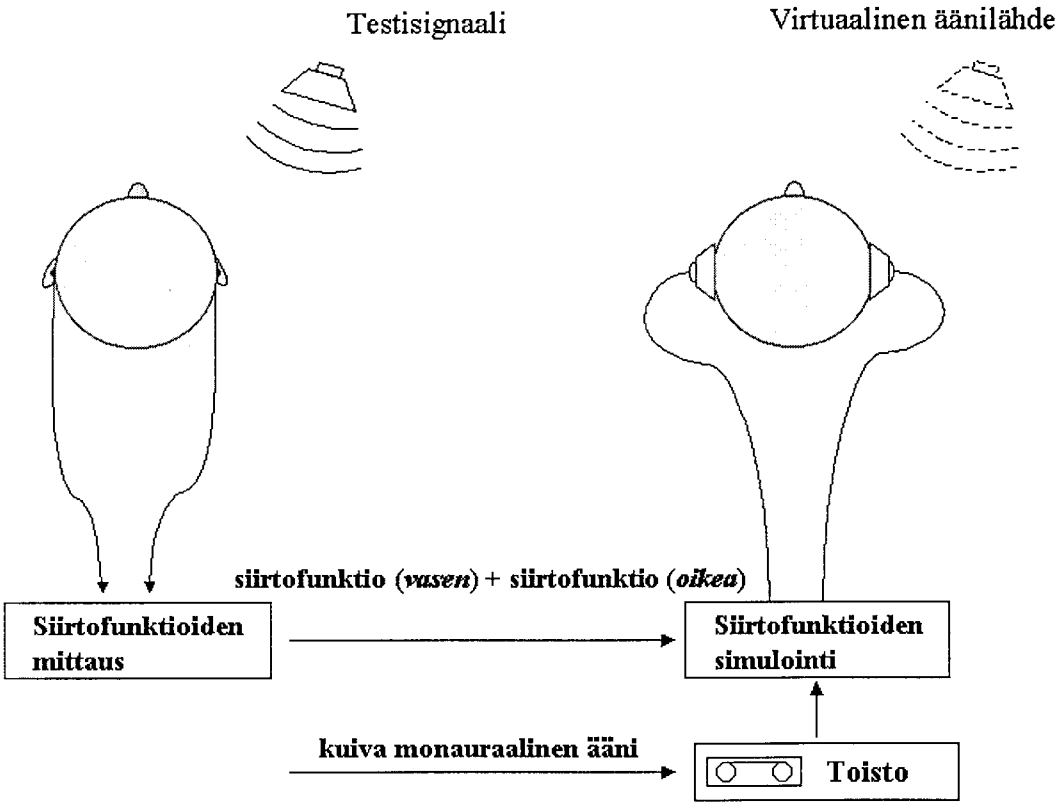
Kuulija voi siis paikantaa ääniä kaikista suunnista oman päänsä ulkopuolelta, myös edestä ja takaa (Wu ym. 1997, 16). Olemme jatkuvasti tällaisen akustisen

kolmiulotteisen äänikentän ympäröimänä. Binauraalinen äänitys ei ole sidottu vertikaaliseen linjaan (toisin kuin stereossa), vaan äänet tuntuvat tulevan tietystä paikasta pään ulkopuolelta. Binauraalinen äänitys kuulostaa realistisemmalta kuin stereoäänitys. Binauraalinen ääni voidaan Jot'n (1997, 56) mukaan toteuttaa kahdella tavalla: (a) äänittämällä äänitapahtuma omassa akustisessa ympäristössään tai (b) syntetisoimalla virtuaalinen äänitapahtuma. KUVIOSSA 2 on havainnollistettu binauraalisen äänen nauhoitusta ja toistoa (kohta (a)) (McGrath 1995, 3), ja KUVIOSSA 3 siirtofunktioiden mittausta ja simulointia (kohta (b)) (McGrath 1995, 3).



KUVIO 2. Binauraalinen äänitys ja toisto

Ensiksi mainitussa tapauksessa äänitys tehdään joko lähimikityksellä tai keinopään (dummy head, artificial head) tai koehenkilön korvakäytäviin asennetuilla mikrofoneilla. Tällöin äänitapahtuma sisältää sekä paikantamiseen että huonekaikuun liittyvät elementit. Menettely kuitenkin hankaloittaa äänen jälkikäsittelyä. Kun taas äänitapahtuma on synteettinen, paikantamiseen ja tilan akustisiin ominaisuuksiin liittyvät parametrit lisätään ääneen signaalinprosessoinnin avulla keinotekoisesti, äänittämisen jälkeen siirtofunktion avulla (Kendall 1995). Äänen binauraalisen käsittelyn avulla voidaan siis luoda kolmiulotteinen äänikenttä kahden kanavan avulla (Jot 1997, 60).



KUVIO 3. Siirtofunktioiden mittaus ja simulointi

Burgess (1992b, 55) kuvaa spatialisointiprosessia seuraavassa esitettävällä tavalla: Kutakin äänilähteen koordinaattia pysty- ja vaakatasolla vastaa yksi siirtofunktio-suodatinpari kumpaakin korvaa varten. Jotta kuulija voi sijoittaa äänen tiettyyn paikkaan äänikentässä, kyseinen ääni on prosessoitava vastaavalla suodatinparilla. Burgess (1992b, 55) esittää silmukka-algoritmin äänten reaaliaikaiselle spatialisoinnille:

Prosessi 1: päivitä äänen sijainti
hanki halutut vaaka- ja pystykoordinaatit
etsi lähimmät siirtofunktio-suodatinparit
hanki haluttu suodatinpari
lähetä suodatinpari prosessi 2:seen

Prosessi 2: käytä suodatinta
jaa monauraalinen lähdesignaali binauraaliseksi
lisää vasen ja oikea suodatin signaaleihin
konvertoi äänikanavat analogisiksi ja soita ääni
korvaa nykyinen suodatinpari uudella (takaisin prosessiin 1)

Burgessin (1992b, 53) mukaan uudet digitaaliset signaaliprosessorit mahdollistavat jo reaaliaikaisen spatialisoinnin. Prosessissa 2 jokainen yksittäinen monauraalinen ääni siis jaetaan oikealle ja vasemmalle kanavalle, jonka jälkeen saadut stereoparit prosessoidaan halutulla suodattimella. Lopuksi kaikki valmiit stereoparit summataan yhdeksi stereotulosteeksi, joka toistetaan joko kuulokkeilla tai kaiuttimilla. Prosessi vaatii koneelta erittäin paljon laskutehoa, varsinkin jos tarkoituksena on luoda monipuolinen useista yksittäisistä äänilähteistä muodostunut kolmiulotteinen äänikenttä.

Alten (1999, 417) mainitsee stereoäänen ja kolmiulotteisen (binauraalisen) äänen lisäksi vielä *surround-äänen* menetelmänä luoda spatiaalinen äänikenttä. Kolmiulotteinen surround-ääni saadaan aikaan sijoittamalla kaiuttimia eri puolille kuulijaa. Tällaiseen monikaiutinjärjestelmään ei kuitenkaan tässä tutkimuksessa puututa, vaan spatialisoidun äänen toisto rajoitetaan tapahtuvaksi

kuulokkeilla tai vaihtoehtoisesti kahdella kaiuttimella (ks. kohta 3.5). Spatialisointi on siis menettelytapa, jolla luodaan kolmiulotteinen äänikenttä, josta voidaan erottaa ja paikantaa äänilähteitä eri suunnissa. Kun äänet on spatialisoitu, kuulija voi erottaa äänikentästä eri lähteitä, keskittyä niihin ja sivuuttaa muut.

Asia erikseen on "mentaalinen" spatialisaatio. Chionin (1990, 69) mukaan perinteisessä monauraalisessa elokuvassa sijainti, josta ääni fyysisesti tulee, ei ole sama kuin mistä se koetaan tulevan. Esimerkiksi jos henkilö elokuvassa kävelee ruudun halki, askeleet koetaan kuuluvan ja seuraavan henkilöä, vaikka ne todellisuudessa kuuluvat samasta staattisesta kaiuttimesta. (Chion 1990, 69.) Tällöin spatiaalisuus koetaan äänen ja kuvan yhteisvaikutuksena mielen tasolla, ei fyysisesti (ks. audiovisuaalinen sopimus, kohta 2.4).

3.4 Auralisaatio

Auralisaatio on visualisaation vastine kuuloaistin alueella (McGrath 1995, 1). Begaultin (1994, 145) mukaan auralisaatiossa luodaan jokin akustinen virtuaalitila huonekaiun mallintamisen ja synteettisen spatialisoinnin avulla. Siirtofunktio-suodattimilla prosessoidaan sekä suora ääni että heijastuneet äänet. Kendallin (1995) mukaan etäisyys- ja tilavaikutelma luodaan ääneen kontrolloimalla ensiheijastumien ja jälkikaiunnan määrää. Kleiner ym. (1993, 861) ovat määritelleet auralisaation seuraavasti:

> Auralization is the process of rendering audible, by physical or mathematical modeling, the sound field of a source in a space, in such a way as to simulate the binaural listening experience at a given position in the modeled space.

Tavoitteena auralisaatiossa on siis luoda vaikutelma jonkin tilan akustisista ominaispiirteistä, siihen liittyvistä äänistä sekä niiden sijainnista äänikentässä riippuen kuulijan sen hetkisestä sijainnista kyseisessä tilassa. Tämä tila voi olla

joko jonkin olemassaolevan paikan virtuaalinen vastine tai täysin kuviteltu tila. Äänimateriaali suodatetaan ja prosessoidaan digitaalisesti, jolloin tuloksena on edellämainittu illuusio, joka on luotu tietyn paikan akustisten tunnuspiirteiden avulla. (Kleiner ym. 1993, 861.) Auralisaatiojärjestelmä koostuu yksinkertaisimmillaan äänilähteestä, tilasta ja kuulijasta, jotka on mallinnettava erikseen (Kleiner ym. 1993, 862; Begault 1994, 4). Siinä missä spatialisoinnilla tarkoitetaan äänen teknistä prosessointimenetelmää, auralisaatiolla viitataan tässä tutkimuksessa kattavampaan prosessiin, joka pitää sisällään koko äänimaiseman mallintamisen. Käyttöliittymässä oleellista on, että mallinnus tähtää tehokkaaseen vuorovaikutukseen ja tavoitteelliseen toimintaan.

3.5 Kaiutin- ja kuulokeäänentoiston vastakkainasettelu

Ongelmana kaiutinäänentoistossa on, että vasemmasta kaiuttimesta kuuluva ääni havaitaan myös oikeassa korvassa ja toisinpäin (crosstalk) (Huopaniemi 1999, 30). Ongelmaa on pyritty minimoimaan kehittämällä tekniikoita ristikkäisten signaalien eliminoimiseksi (crosstalk cancelled binaural processing). Huopaniemi (1999, 30) mainitsee kaksi rajoitusta binauraalisessa kaiutinkuuntelussa: ensinnäkin kuulijan on oltava paikallaan ja toiseksi kuunteluhuoneen on oltava mahdollisimman kaiuton. Kendallin (1995) mukaan 1 millisekunnin aikana tulevat heijastukset tuhoavat siirtofunktion vaikutuksen, joten heijastukset kaiuttimien ja kuulijan lähellä on eliminoitava.

Kolmiulotteinen spatialisointi kaiuttimilla toistettaessa onnistuu siis parhaiten silloin, kun kuulijan sijainti kaiuttimiin nähden on etukäteen tiedossa ja pysyy samana koko ajan. Tämä viittaisi siihen, että kolmiulotteista ääntä voitaisiin käyttää kotitietokoneeseen liitettyjen kahden kaiuttimen avulla, koska tietokoneen käyttäjä istuu koko ajan paikallaan. Ongelmana ovat kuitenkin suuret tehovaatimukset ja kaiun eliminointi kotioloissa. "Crosstalkin" luotettava eliminointi voisi lisätä kaksikanavaisen, kolmiuloitteisen äänentoiston

mahdollisuuksia multimediakäyttöliittymissä ja kotikoneissa (Jot 1997, 62). Kaiutinkuuntelu on helpompi toteuttaa kuin kuulokekuuntelu, mutta se ei tuota yhtä luotettavaa tulosta. Kaiutinkuuntelun onnistuminen riippuu Jot'n (1997, 60) mukaan ainakin käytetystä tekniikasta, simuloitavien äänten suunnista, kaiuttimien suuntaavuudesta sekä kuunteluhuoneen koosta ja akustisista ominaisuuksista. Kaiutinkuuntelussa paikantaminen etualalla on huomattavasti helpompaa kuin paikantaminen takana tai yläpuolella (Jot 1997, 62).

Etuna kuulokeäänentoistossa taas on, että kuunteluhuoneen akustiikka tai kuulijan sijainti huoneessa eivät vaikuta kuulokokemukseen (Huopaniemi 1999, 30). Kuulokkeet myös mahdollistavat spatiaalisen äänilähteen sijainnin tarkemman kontrollin (Wenzel ym. 1991, 351; McGrath 1995, 4; Begault 1999, 14). Kendallin (1995) mielestä haittana voi kuitenkin olla se, että huolimatta vaihe- ja voimakkuuseroista ääntä ei koeta tulevan riittävästi pään ulkopuolelta, vaan ääni vain lateralisoituu joko vasempaan tai oikeaan kuulokkeeseen. Toinen ongelma on etu- ja taka-alan sekoittaminen, jota kuitenkin voidaan vähentää liioittelemalla huomattavasti niitä äänen spektraalisia muunnoksia, joita edessä ja takana olevat äänilähteet aiheuttavat (Kendall 1995). Kuulokekuuntelun avulla voidaan luoda luonnollisen tuntuinen kolmiulotteinen äänikenttä, mutta haittana on kallis ja monimutkainen toteutus. Tekniikan kehittyessä tämä ongelma poistunee, jolloin myös kaupallisissa sovelluksissa voidaan kuulla hyvätasoista spatiaalista ääntä.

3.6 Äänen spatialisoinnin ongelmat

Suurimmat haasteet kolmiulotteisen äänikentän toteuttamisessa ovat Begaultin (1991, 865) ja Kleinerin ym. (1993, 865) mukaan seuraavat: edessä ja takana olevien peilikuvaäänten poistaminen, paikantamisvirheiden minimointi, ihmisen havaintokyvylle riittävän siirtofunktiomittaustason saavuttaminen

mahdollisimman pienellä datamäärällä sekä ratkaisun löytäminen mitattujen siirtofunktioiden ja havaitun kuulokuvan välillä vallitseviin ristiriitoihin.

Kuulijan voi olla vaikea paikantaa ääni edestä tai takaa tulevaksi (Burgess 1992b, 56). Kendallin (1995) mukaan nämä paikantamisvaikeudet pystyakselilla johtuvat korvien saamasta yhtäsuuresta vaihe- ja voimakkuusinformaatiosta, eli peilikuvaäänestä. Tällöin paikantaminen perustuu vähäisiin spektraalisiin eroihin siirtofunktion perusteella (Begault 1991, 866; Jauhiainen 1995, 152; Kendall 1995). Burgessin (1992b, 57) mukaan pääasialliset vihjeet äänen sijainnista kuulijan edessä tai takana saadaan pään kääntämisen ja korvalehtien vaikutuksen perusteella. Myös simuloidun huoneen ensiheijastumien lisääminen spatialisoituun ääneen helpottaa Burgessin (1992b, 57) mukaan äänen paikantumista eteen tai taakse.

Näitä paikantamiseen liittyviä ongelmia voidaan huomattavasti vähentää, jos spatialisaatiojärjestelmään liitetään erityinen laite, joka tarkkailee pään asentoa (McGrath 1995, 5). Tällainen "head tracker" välittää tiedot pään asennosta tietokoneelle, joka päivittää samanaikaisesti suodatinparien suuntatiedot. Näin kuulija kokee päätä kääntäessään äänilähteen pysyvän koko ajan samassa paikassa (Kendall 1995). Tällä tavalla voidaan todellisuutta jäljentämällä lisätä kokemusta, että ääni tulee pään ulkopuolelta. Jos ääntä ei koeta tulevan riittävän vahvasti pään ulkopuolelta, se saattaa johtua epätarkasti mitatusta siirtofunktiosta (Burgess 1992b, 56).

Synteettisesti spatialisoitu äänikenttä - virtuaalinen kuulokokemus - voi poiketa suurestikin jäljiteltävästä akustisesta kuulokokemuksesta. Virheitä voi esiintyä joko äänilähteen, tilan tai kuuntelijan mallinnuksessa. Kolme keskeisintä paikantamisvirheiden aiheuttajaa ovat epätarkat siirtofunktiot, käytettävien äänten erilaiset ominaispiirteet (esim. taajuus ja voimakkuus) sekä järjestelmän spatiaalinen erotuskyky (Begault 1991, 865). Begault (1991, 865) toteaa myös,

että vaikka nämä kaikki kolme tekijää olisivatkin kunnossa, ongelmia voi silti aiheuttaa kuulijoiden vaihteleva paikantamiskyky. Useimmiten on tyydyttävä kompromissiin ja käytettävä keskimääräisiä siirtofunktioita. Lukemattomien yksittäisten äänten spatialisaatio eri puolille kolmiuloitteista äänikenttää vaatii tietokoneelta suurta suorituskykyä. Päämääränä spatialisoinnissa on pidettävä riittävän tarkkaa akustisen ympäristön virtuaalitoistoa mahdollisimman pienellä datamäärällä. Akustista tilaa ei tietenkään aina tarvitse edes täyttää. Kuten Chion (1990, 154) sanoo: spatiaalisuus suurentaa sitä tilaa, joka mahdollisesti voidaan täyttää äänillä. Tällöin päästään Blyn (1982, 375) mainitsemaan uuteen ulottuvuuteen. Seuraavassa kohdassa käsitellään tätä uutta ulottuvuutta muutaman sovelluksen kautta.

3.7 Spatiaalista ääntä hyödyntäviä sovelluksia

Akustinen ekologia sopii hyvin lähtökohdaksi käyttöliittymäsuunnittelulle, koska spatiaalinen ääni kuuluu luonnolliseen ääniympäristöömme. Käyttöliittymäsuunnittelija ei kuitenkaan ole sidottu fyysiseen ääniympäristöön (Kramer 1994, 56). Tietokonepeleissä on käytetty jonkin verran kolmiulotteista ääntä, joskin Goosen ja Möllerin (1999, 366) mukaan näitä ääniefektejä on käytetty lähes yksinomaan tunnelman luomiseen, eikä peleissä ole juurikaan hyödynnetty äänten spatialisointia. Spatialisoinnin avulla voidaan kuitenkin tarjota lisäinformaatiota ja lisätä pelin todentuntuisuutta. Wenzelin ym. (1991, 351) ja Wenzelin (1994, 38) mukaan spatiaalisen äänen käytöllä on ainakin kaksi etua: (a) objektien ja niiden välisten suhteiden havaitseminen kolmiulotteisessa tehtäväavaruudessa ja (b) useiden äänivirtojen havaitseminen samanaikaisesti, jolloin kuulija voi vaihdella huomiotaan eri äänivirtojen kesken. Tällöin siis yhdistetään objektien merkityksiä ja suuntatietoa. Tätä seikkaa on käytetty hyväksi jo useissa sovelluksissa ja prototyypeissä, joista seuraavassa esitellään tärkeimmät.

"AudioStreamer"-sovellus perustuu äänen spatiaaliselle erottelulle ja samanaikaiselle kuuntelulle; käyttäjä voi päätään kääntämällä valita mieleisensä kanavan kolmesta vaihtoehdosta, jolloin kyseisen kanavan äänenvoimakkuus kasvaa (Schmandt & Mullins 1995, 218). Sovelluksessa käyttäjä siis kommunikoi käyttöliittymän kanssa pään liikkeiden avulla (manuaalisen käyttöliittymän sijaan), ja peräkkäinen äänentoisto on korvattu monikanavaisella äänentoistolla. Sovellus perustuu paikallaan pysyville äänilähteille. Sawhney ja Schmandt (1997, 110) ovat käyttäneet radiota metaforana äänitiedostojen selaamisessa ja kokeilleet ajatuksiaan "Nomadic Radio" -prototyypissä. He esittelevät kolme spatiaalisen kuuntelun toimintatilaa (Sawhney & Schmandt 1997, 111):

- **Lähetys:** viesti lähetetään tietystä sijainnista tilassa. Viesti on kuultavissa taka-alalla hetken ennen häipymistään kuulumattomiin, ellei käyttäjä reagoi siihen ennen sitä. Vertaus perustuu radion kuunteluun, jossa kuulija aktivoituu vasta sitten, kun kuulee mielenkiintoisen kohdan.
- **Selaus:** käyttäjä voi selata viestejä ja tuoda jonkun niistä etualalle. Tämä vertaus perustuu radiokanavien aktiiviseen kuunteluun ja selailuun, kunnes mieluisa kanava löytyy.
- **Vilkaiseminen:** Nopea katsaus kaikkiin viesteihin siirtämällä ne kuuntelualueen keskelle.

Radio-metafora vaikuttaa olevan käyttökelpoinen (vaikka siinä onkin lisäominaisuutena spatiaalisuus). Koska ääni on väliaikaista, sen avulla ei voi selata useita tiedostoja samalla tavalla kuin tekstiä ja kuvia näköaistin avulla (Kobayashi & Schmandt 1997, 194; Sawhney & Schmandt 1997, 109-110). Äänen sijainnin avulla voidaan antaa tietoa viestin kategoriasta tai kiireellisyydestä (Sawhney & Schmandt 1997, 109-110). "Multimedia Browser" -prototyypissä Fernström ja Bannon (1997) kokeilivat sijoittaa useita äänitiedostoja stereoäänikenttään niiden visuaalisen sijainnin mukaan näytöllä.

Osoittamalla nuolella tiettyä visuaalista objektia, vastaava äänitiedosto soi samassa sijainnissa.

Kobayashi ja Schmandt (1997) kehittivät spatiaalisen käyttöliittymän ("Dynamic Soundscape"), joka perustuu äänen sijainnin muistamiselle ajassa. Eteen- ja takaisinkelauksen sijaan kuuntelija vaihtaa huomiotaan liikkuvien äänilähteiden kesken (jotka toistavat saman äänityksen eri kohtia samanaikaisesti), jolloin kuuntelija voi kytkeä tietyn ajankohdan äänitiedostosta tiettyyn sijaintiin (Kobayashi & Schmandt 1997, 194). Käyttäjä voi siten selata äänitapahtumia samoin kuin selaisi visuaalisia objekteja. Vaikka käyttäjä keskittyy yhteen äänitapahtumaan kerrallaan, hän kuulee muutkin äänitapahtumat taustalla. Sen sijaan, että käyttäjät ajattelisivat "aihe jonka kuulin 20 sekuntia sitten", he voivat ajatella "aihe jonka kuulin takavasemmalta". Tällä tavalla spatiaalisen muistin avulla korvataan ajallisen muistin heikkouksia. (Kobayashi & Schmandt 1997, 196.)

Toteutuksessa oli kuitenkin muutamia ongelmia: kuuntelijoilla oli vaikeuksia muistaa äänityksen tietyn kohdan sijainti (Kobayashi & Schmandt 1997, 196-198). Muistamista vaikeutti ainakin se, että äänitapahtumat olivat liikkuvia. Mitä hitaammin äänitapahtumat liikkuivat, sitä paremmin kuuntelijat kytkivät kohdan sijaintiin. Toinen ongelma oli, että spatiaalisen muistin resoluutio on huono: voimme sanoa, että "kohde vasemmalla ylhäällä", mutta emme että "kohde 38 astetta keskikohdasta vasemmalle". Spatiaalinen sijainti voidaan muistaa vain ympyrän kahdestoistaosan tarkkuudella. (Kobayashi & Schmandt 1997, 196-198.)

Tekijöiden idea on kuitenkin kehityskelpoinen: vastakkainasettelu (peräkkäisen) eteen- ja takaisinkelauksen ja selektiivisen (samanaikaisen) kuuntelun välillä tuntuu toimivan. Spatiaalinen esitys tarjoaa mahdollisuuden järjestellä informaatiota ja hyödyntää käyttäjän muistia paremmin. Goose ja Möller (1999,

363) taas ovat esittäneet käsitteellisen mallin HTML-dokumentin rakenteen kytkemiselle spatiaaliseen ääniavaruuteen. Mallissa on uutta se, että siinä pyritään ensi kertaa kytkemään dokumentin rakenne spatiaaliseen ääneen, tavanomaisen dokumentin sisällön sijaan. Walker ja Brewster (2000) ovat tutkineet spatiaalisen äänen käyttöä mobiileissa laitteissa ja kokeilleet spatialisoitua ääntä ja sijainnin muuttumista kuvaamaan tiedoston latautumista. He päättelivät, että spatiaalinen ääni lisäsi tarkkuutta taustaprosessin monitoroinnissa ja tehosti keskittymistä samanaikaisesti suoritettuun päätehtävään (Walker & Brewster 2000, 12).

Edellä esitetyn perusteella voidaan päätellä, että spatiaalinen ääni muodostaa mielenkiintoisen uuden ulottuvuuden käyttöliittymään, varsinkin yhdistettäessä visuaaliseen informaatioon. Luvussa on esitetty paikantamiseen vaikuttavat tekijät, korvan "hämäämiseen" perustuva synteettisen spatialisoinnin periaate sekä keskeisimmät ongelmat sen toteutuksessa. Sovelluksista voidaan kuitenkin päätellä, että vaikeudet ovat voitettavissa: 1990-luvun loppuun mennessä jo useat sovellukset ovat menestyksellisesti hyödyntäneet spatiaalista ääntä. Mikä parasta, nämä sovellukset ovat osoittaneet käytännön toteutuksen ja koehenkilöiden kokemuksien kautta ne kohdat, joissa on parantamisen varaa. Kehitettävää todellakin on, mutta tekniikan edistyessä ongelmaksi jää edelleen se, miten spatiaalista ääntä tulisi käyttää.

Spatiaalinen ääni muodostaa tavallaan kannettavan "kuplan" käyttäjän pään ympärille, jolloin informaatiota voidaan järjestellä laajemmalle alueelle kuin visuaalista informaatiota. Tämä spatiaalinen kupla voidaan jakaa pienemmiksi tiloiksi, jotka sisältävät paikallisia ja globaaleja merkityksellisiä ääniä. Enkoodaamalla äänen parametri (tässä tapauksessa spatiaalinen ääni) kuhunkin ajassa olevaan pisteeseen, saadaan aikaan "laulu", jota kuuntelemalla voidaan määritellä merkityksellisiä kohtia ja muutoksia (Bly 1982, 374). Tämä on seuraavan luvun aihe.

4 ÄÄNIMAISEMA-ANALYYSI KÄYTTÖLIITTYMÄSSÄ

Tässä luvussa pyritään yhdistämään äänen suunnan tuottama merkityssisältö käyttöliittymän toimintoihin. Luvussa tarkastellaan äänen mahdollisia funktioita käyttöliittymässä, jonka jälkeen analysoidaan ääniä semioottisessa viitekehyksessä. Lopuksi tarkastellaan erilaisia äänten luokittelujärjestelmiä. Ensin on kuitenkin pohdittava hiljaisuuden merkitystä.

4.1 Hiljaisuus

Kun näin kauan olemme puhuneet äänestä, on tarpeen käsitellä ei-ääntä: hiljaisuutta. Chion (1990, 57) korostaa, että hiljaisuus ei ole koskaan neutraalia tyhjyyttä; se on aiemmin kuullun tai odotetun, kuvitellun äänen negatiivi. Kun hiljaisuus ennakoi ääntä, hermostunut odotus tekee sen eläväksi; kun taas hiljaisuus keskeyttää äänen tai seuraa sitä, hiljaisuus on täynnä sitä edeltäneen äänen jälkivaikutusta niin kauan, kuin muisti jaksaa sitä sellaisena pitää (Schafer 1977, 257). Näin ääni siis luo kontrasteja. Länsimaisessa kirjallisuudessa hiljaisuus on kuitenkin usein kuvattu ahdistavaksi, yksinäiseksi tai raskaaksi, vain harvoin täyttymyksen tai tyytyväisyyden tilaksi tai muuten positiiviseksi ilmiöksi (Schafer 1977, 146). Schaferin (1977, 256) mielestä länsimaiselle ihmiselle hiljaisuus on negatiivista; se pitää rikkoa vaikka väkisin. Toisaalta absoluuttista hiljaisuutta ei ole olemassakaan, koska aina on jotain, joka aiheuttaa ääntä.

Gaverin "SonicFinder" oli menestys: totuttuaan siihen käyttäjät valittivat äänten puuttumista käyttäessään äänetöntä versiota (1989, 82). Äänikäyttöliittymään voi siis tottua. Onko meidät kuitenkin pakotettu tottumaan äänettömään käyttöliittymään ja sen piinaavaan, yksinäiseen hiljaisuuteen; hiljaisuuteen joka on vain tyhjä negatiivi ilman informaatioarvoa? Kun käyttöliittymä jaetaan äänettömyyteen ja kuuluvuuteen, myös hiljaisuus saa uuden merkityksen. Jälleen

päästään kontekstiin: hiljaisuus merkitsee eri asioita eri konteksteissa. Oleellista on se, että tehdään selvä ero hiljaisuuden ja äänen välillä. On aika olla hiljaa ja aika olla äänekäs.

Buxtonin (1989) mukaan ihmiset saattavat vastustaa äänen käyttöä käyttöliittymässä perustellen sitä seuraavasti: "Työskentelen hälyisässä toimistossa, enkä halua enää enempää melua häiritsemään työtäni" (Cohen 1994, 513). Buxton vastaa sanomalla, että ääni on jo muutenkin läsnä kaikkialla ja että jos kontrolloisimme paremmin ääniympäristöämme, saisimme kokonaisäänimaiseman vähemmän vastenmieliseksi. Ihmiset suhtautuvat siis ääneen kaksijakoisesti: joko vastustavat sitä voimakkaasti tai suhtautuvat todella innokkaasti. Sitä paitsi meluisassa toimistossakin voi käyttää kuulokkeita, jotka eivät sulje pois toimiston muita tärkeitä ääniä. Brewster (1994, 6) korostaa, että jos äänet ovat informatiivisia, käyttäjät eivät halua laittaa niitä pois päältä. Onkin mielenkiintoista, että äänen osuutta käyttöliittymässä kuvataan usein adjektiivilla "häiritsevä". Jos käyttöliittymän äänet ovat huonosti suunniteltuja, niitä saatetaan pitää myös "huvittavina". Tällaiset määritelmät osoittavat, että paljon on vielä tekemättä sekä asenteiden että tekniikan osalta. Harvemmin graafisia elementtejä pidetään häiritsevinä. Selvää kuitenkin on, että visuaalisesti liian täynnä oleva näyttö ei ole lainkaan miellyttävä käyttää. Eikö tämäkin ole häiritsevää?

Vaikka ääntä käyttöliittymässä saatetaan vastustaa (ja vaikka ääni on käytännössä suunniteltu käyttöliittymistä pois!), ihmiset joka tapauksessa kuuntelevat tietokonettaan saadakseen tietoa prosesseista, joita he eivät voi nähdä. Esimerkkinä tästä on levyaseman hurina tai tulostimen ääni. (Gaver 1989, 70; Walker & Brewster 2000, 2.) Toisaalta suhtautuminen ääneen on ymmärrettävää: emme voi sulkea korviamme (toisin kuin silmiämme), joten kuuloaistin ainoa suojautumiskeino on epämieluisten äänten suodattaminen ja keskittyminen miellyttävimpiin ääniin (Schafer 1977, 11). On kuitenkin otettava huomioon, että

myös epämiellyttävät äänet ovat merkitseviä. Seuraavaksi tarkastellaan äänen funktioita käyttöliittymässä.

4.2 Äänen käyttötarkoitus

Ääntä on yleensä käytetty elokuvassa ainakin kuvien yhdistämiseen (ylimeno) tai ilmapiirin luomiseen (Chion 1990, 47). Perinteisessä elokuvatutkimuksessa elokuva ilman ääntä pysyy elokuvana, mutta elokuva ilman kuvaa ei ole enää elokuva (Chion 1990, 143). Tämä ajattelumalli kuvaa hyvin visuaalisen informaation ylivaltaa, vaikka kokeellisissa elokuvissa näitä kirjoittamattomia sääntöjä onkin pyritty tietoisesti rikkomaan. Chionin (1990, 145-148) mukaan äänielokuvan alkuaikoina ei ollut teknisesti mahdollista käyttää monia ääniä päällekkäin, koska silloin ne eivät olisi erottuneet toisistaan. Jos käytettiin useita ääniä, yhden piti olla ylitse muiden. Tähän saattoi Chionin mielestä olla myös kulttuurisia syitä: melu ei useimpien ihmisten mielestä ole esteettisesti kiehtovaa. (Chion 1990, 145-148.) Melu on määritelty epämieluisaksi, epämusikaaliseksi tai kovaksi ääneksi (vaikka "epämieluisa ääni" on subjektiivinen kokemus, se nojaa myös yhteisön tuottamiin sopimuksiin) (Schafer 1977, 182-183).

Schafer (1977, 205) pitää maailman äänimaisemaa yhtenä suurena sävellyksenä, johon kaikki ihmiset osallistuvat säveltäjinä, esiintyjinä ja kuulijoina. Prosessiin kuuluu eri äänten analysointi ja päättäminen siitä, mitä ääniä haluamme säilyttää ja mitä poistaa. Tällä tavalla voidaan päästä melusta eroon. Vaikka tällainen näkemys on melko naivi, se saattaa olla mahdollista toteuttaa pienemmässä mittakaavassa: käyttöliittymässä suunnittelija voi toimia säveltäjänä, esiintyjänä ja kuuntelijana.

Chion (1990, 73) esittelee mielenkiintoisen viitekehyksen äänen kolmijakoisesta roolista elokuvassa: äänitapahtuma voi olla joko (a) näytöllä (onscreen), (b) näytön ulkopuolella näkymättömissä (offscreen) tai (c) näkymättömissä, mutta ei

myöskään missään suhteessa näytöllä oleviin tapahtumiin (nondiegetic, esimerkiksi musiikki). Chion kritisoi itsekin tätä jakoa puutteelliseksi ja lisää (1990, 82-83), että jos suljemme silmämme tai katsomme muualle, näytön ulkopuoliset äänet muuttuvat näytöllä oleviksi ääniksi. Äänen rooli syntyy siis visuaalisen ja äänellisen yhteistyönä. Chion (1990, 75) lisää tähän kolmijakoon vielä äänitapahtumat, jotka ympäröivät tilaa ilman, että kiinnitetään huomiota niiden tunnistamiseen tai tarpeeseen nähdä niiden lähde. Niiden avulla voitaisiin kuitenkin tunnistaa jokin tila. Chionin näkökulma lähenee tässä kohdin Schaferin (1977) ja Sauen (2000) näkemyksiä äänimaisemasta. Chionin elokuvateoriat ovat erityisen hyödyllisiä myös käyttöliittymäkontekstissa, koska jako näytöllä ja sen ulkopuolella oleviin äänitapahtumiin viittaa piiloinformaation olemassaoloon.

Äänen funktiona ei ole (pelkästään) viihdyttäminen, vaan piiloinformaation tunnistaminen ja esittäminen käyttäjälle (Gaver 1993b, 228; Blattner ym. 1994, 467). Tämä näkemys kiteyttää oleellisen ja pitää sisällään laajan kirjon mahdollisia käyttötarkoituksia. Herefordin ja Winnin mukaan (1994, 212-213) äänellä on käyttöliittymässä kaksi funktiota: (a) ääni kertoo käyttäjälle järjestelmän tilasta kullakin hetkellä hälytysäänillä tai "tarkkailemalla" taustalla kunnes käyttäjä vaatii tietoa järjestelmän tilasta ja (b) välittää ohjelmien tuottamaa tietoa. Buxton (1990, 124-127) taas jakaa äänet käyttöliittymässä kolmeen luokkaan sen mukaan, minkälaista informaatiota ne välittävät:

- hälytys- ja varoitusäänet
- järjestelmän tilasta kertovat äänet
- koodatut viestit.

Varoitusäänten (yleensä kovia ja korkeita ääniä) tarkoituksena on keskeyttää meneillään oleva tehtävä. Järjestelmän tilasta kertovien äänten tarkoituksena on välittää informaatiota meneillään olevasta tehtävästä tai prosessista. Tällaiset äänet ovat yleensä pitkiä ääniä tai toistuvia kuvioita, jotka häipyvät taka-alalle

prosessin päättymisen jälkeen. Näin käyttäjä voi keskittyä päätehtävään (kunnes ääni taas ilmaantuu etualalle ilmoittaen prosessissa tapahtuvasta muutoksesta). Ihminen ei pysy kauaa tietoisena staattisista äänistä, mutta aktivoituu heti, kun äänessä tapahtuu jokin muutos (esimerkiksi jos autolla ajaessa moottorin ääni yhtäkkiä muuttuu). Ihminen pystyy tarkkailemaan useita samanaikaisia ääniä taka-alalla (mikäli äänet ovat riittävän erilaisia), mutta reagoimaan vain yhteen tai kahteen samanaikaisesti. Koodattujen viestien tarkoituksena taas on välittää kvantitatiivista informaatiota. Tällaiset äänet ovat vaihtelevan monimuotoisia, toisin kuin hälytysäänet ja järjestelmän tilasta kertovat äänet. (Buxton 1990, 124-127.)

Beaudouin-Lafon ja Gaver (1994, 49-50) vuorostaan jakavat äänen funktiot seuraavasti: ääni (a) antaa palautetta käyttäjän toiminnoista, (b) ilmoittaa järjestelmän toiminnasta ja (c) lisää tietoisuutta muiden käyttäjien toiminnoista. Esimerkkinä kohdasta (a) voidaan mainita Gaverin "SonicFinder" (1989) ja kohdasta (b) Gaverin, Smithin ja O'Shean "ARKOLA simulation" (1991). Kohtaa (c) ei tässä käsitellä, koska kiinnostuksen kohteena on yksittäinen käyttäjä. Edelliseen kolmijakoon voitaisiin lisätä vielä tieto eri äänitapahtumien sijainnista ja merkityksistä tilassa, mistä päästään tämän tutkimuksen aiheeseen. Toisaalta spatiaalinen ääni ei ole oma kohtansa; pikemminkin se on uutena ulottuvuutena läsnä kaikissa niissä funktioissa, joissa ääni voi käyttöliittymässä toimia.

Käytännön esimerkkinä äänen käytöstä mainittakoon Albersin ja Bergmanin (1995, 318) laajennus Mosaic-selaimeen nimeltä "Audible Web" (prototyyppi), jonka tarkoituksena on äänen avulla (a) antaa informaatiota tiedonsiirron etenemisestä, (b) antaa palautetta käyttäjän toimista sekä (c) antaa sisällöllistä palautetta linkkien koostumuksesta navigoinnin helpottamiseksi (esimerkiksi tiedoston tyyppi, koko ja latausaika). Prototyypissä on tekijöiden mukaan pyritty liittämään ääni huomaamattomasti kokonaisuuteen luonnolliseksi osaksi

vuorovaikutusta. Juuri tämä tulisi olla äänen osuus: ei visuaalisuuteen jälkeenpäin liitetty osa, vaan alusta asti mukaan suunniteltu.

Brown ym. (1989, 346) päätyivät tutkimuksessaan siihen, että käyttäjät voivat erottaa äänestä useita samanaikaisia informaatioyksikköjä ja reagoida niihin tarjotun informaation mukaisesti. Tutkimuksesta kävi myös ilmi, että esittämällä informaatiota sekä äänen että grafiikan avulla voidaan näyttöä käyttää useisiin samanaikaisiin tehtäviin. Informaatio voi kuitenkin jäädä käyttöliittymässä piiloon ainakin seuraavista syistä (Brewster, Wright & Edwards 1995, 3):

- Informaatio ei ole saatavilla esimerkiksi näytön pienuuden takia.
- Informaatio on saatavilla, mutta vaikea saada (esimerkiksi tiedoston koko).
- Informaatiota on liikaa (visuaalinen yliannostus).
- Käyttäjän katse on suuntautunut väärään paikkaan.

Viimeksi mainittu kohta on tärkeä. Kuten kohdassa 2.2 todettiin, tapahtumalla on kaksi kohtaa ajassa: itse tapahtuma ja sen havaitseminen. Näiden välinen ero voi olla suuri, ellei käyttäjä huomaa katsoa oikeaan paikkaan. Jos kaikki tämä piiloon jäävä informaatio saadaan äänen avulla esille, lisääntyy käyttöliittymän tehokkuus ratkaisevasti. Ongelma on näin ollen selvä: minkälaisia ääniä pitäisi käyttää? Brewster (1994) pyrkii väitöskirjassaan tutkimaan, minkälaisia ääniä pitäisi käyttää käyttöliittymässä, mutta ottaa silti symboliset äänet annettuina ja sivuuttaa ikoniset äänet ilman perusteluja. Tällainen näkökulma on ollut vallitseva 1990-luvun loppupuolella, jolloin symbolisia musiikillisia ääniä on pyritty liittämään käyttöliittymän toimintoihin. 1980-luvun ja 1990-luvun alun ikoniset kokeilut ja akustinen ekologia on jostain syystä jätetty vähemmälle huomiolle. Seuraavassa esitellään näitä tutkimuksia ja pohditaan sitä, miten spatiaalisen äänen avulla voitaisiin tunnistaa ja esittää piilossa olevaa informaatiota.

4.3 Ikoniset äänet

Gaver (1986, 168) määrittelee *ikoniset äänet* (auditory icons) ääniksi, jotka välittävät informaatiota tietystä tapahtumasta. Oleellista ikonisissa äänissä on se, että ne muistuttavat objektiaan (Tarasti 1990, 29; Fiske 2000, 70): äänellä on siis suora yhteys vastaavaan äänilähteeseen, jolloin ne ovat ikonisessa suhteessa toisiinsa. Gaverin ajattelussa keskeistä on, että ääni välittää informaatiota äänilähteen ja sen ympäristön ominaisuuksista, ei äänestä itsessään. Hänen mukaansa saamme informaatiota ympäröivästä maailmasta äänen avulla (ks. jokapäiväinen kuunteleminen, kohta 2.5). Gaverin (1993b, 228) mukaan liittämällä ikonisiin ääniin parametreja (esimerkiksi koko tai nopeus) voidaan viitata tapahtumien ja objektien kategorioiden lisäksi (visuaalisten ikonien tapaan) myös niiden ulottuvuuksiin. Esimerkiksi tiedoston ollessa iso se myös kuulostaa isolta. Parametrisoitujen ikonisten äänten hyvänä puolena voidaan pitää sitä, että informaatiota saadaan enemmän suoraan havainnoimalla kuin symbolisesti opettelemisen kautta. (Gaver 1993b, 228.)

Brownin ym. (1989, 340) mielestä saattaa olla vaikea löytää luonnollista äänellistä vastinetta objektin spatiaaliselle visuaaliselle sijainnille. Näkemys on melko yksioikoinen: edellä on jo osoitettu, että objekti voidaan sijoittaa sellaiseen sijaintiin äänimaisemassa, joka vastaa objektin visuaalista sijaintia. Toisekseen, objektin äänellisen ja kuvallisen esityksen ei välttämättä edes tarvitse olla yhteneväinen. Oleellista on se, mitä spatiaalinen ääni merkitsee.

Mansurin ym. (1985, 117) mukaan ikonisten äänten avulla voidaan informaatiota välittää nopeammin kuin tekstin tai syntetisoidun puheäänen avulla. Ääni-ikoneita käyttämällä voidaan heidän mukaansa saavuttaa seuraavia hyötyjä:

- Käyttäjä voi rauhassa keskittyä päätehtävään luottaen siihen, että ääni-ikoni ilmoittaa tärkeästä väliin tulevasta tapahtumasta.
- Käyttäjän ei tarvitse katsoa näytölle saadakseen ääni-ikonin välittämän tiedon.
- Hyvin suunnitellut ääni-ikonit voivat olla helpommin opittavissa kuin visuaaliset ikonit.
- Ääni-ikonien avulla voidaan korvata tekstimuodossa olevat ilmoitukset ja siten keventää visuaalista kuormaa.
- Ääntä voidaan käyttää silloin kun näköaistia ei voida käyttää (esim. jos käyttäjä on näkövammainen, käyttäjä ei ole koneen ääressä tai jos käyttäjä on yhteydessä järjestelmään puhelimen välityksellä). (Mansur ym. 1985, 117.)

Gaverin ym. (1991, 85) mukaan äänen attribuutit ja haluttu informaatio ovat usein keinotekoisessa suhteessa toisiinsa. Ne voidaan saattaa lähempään viittaussuhteeseen, jos äänitapahtumien attribuutit esittävät käyttöliittymän tapahtumien attribuutteja. Gaver väittää myös, että ikoniset äänet eivät ole niin ärsyttäviä kuin musiikilliset äänet, koska ikonien avulla voidaan täydentää ja laajentaa jo olemassaolevaa ääniympäristöä. Suunniteltaessa ääniä käyttöliittymään on tärkeää, että käytetään ääniä tietyssä funktiossa, ikoni viittaa mahdollisimman tarkasti objektiin (äänilähteeseen) ja että äänet ovat selkeästi eroteltavissa toisistaan. Käyttöliittymän akustinen ekologia tulisi olla kuultavissa sekä kokonaisääniympäristönä että erillisinä äänivirtoina, jotka välittävät informaatiota yksittäisistä objekteista. (Gaver ym. 1991, 85.)

Gaver (1986, 168-169) antaa esimerkin ikonisesta äänestä viestin saapuessa sähköpostijärjestelmään: kirje putoaa kirjelaatikkoon, mistä kuuluu tunnistettava ääni. Jos kirje on iso, kuuluu "painava" ääni. Paperimainen ääni ilmaisee, että kyseessä on tekstitiedosto. Ääni tulee vaimeana vasemmalta: kirjelaatikko on siis oltava näytön vasemmalla puolella ja toisen ikkunan takana. Kaiku (suuri tyhjä huone) ilmaisee, että järjestelmä ei ole liian kuormitettu. Äänet siis välittävät

informaatiota materiaaleista, jotka ovat toistensa kanssa vuorovaikutuksessa tietyssä sijainnissa tietyssä ympäristössä. (Gaver 1986, 168-169.) Gaver (1986, 173) painottaa kuitenkin, ettei ääni-ikonien tarvitse olla täysin realistisia, riittää että ne esittävät oleellisimmat piirteet äänitapahtumasta.

Erilaiset vuorovaikutukset ääniobjektien välillä välittävät erilaista tietoa materiaaleista: esimerkiksi isku antaa informaatiota objektin koosta, kun taas raapaisu vihjaa enemmän objektin koostumukseen (Gaver 1986, 173). Vanderveer (1979) toisti 30 erilaista ääntä nauhalta (esimerkiksi käsien taputus tai paperin repiminen) ja pyysi koehenkilöitä tunnistamaan äänet. Vanderveerin tutkimus osoitti, että koehenkilöt tunnistivat äänilähteet sekä tapahtumat jotka aiheuttivat ne, ja turvautuivat kuvaamaan äänen fyysisiä attribuutteja vain mikäli eivät tunnistaneet äänilähdettä. (Gaver 1989, 72.) Vanderveerin tutkimus tukee näkemystä, että ihmiset kuuntelevat saadakseen tietoa äänilähteistä, harvemmin ääntä itsessään.

Gaver (1989, 77) määrittelee käyttöliittymän tapahtumat toiminnoiksi, jotka kohdistuvat kappaleisiin. Kappaleet vuorostaan koostuvat objekteista, joihin kuuluu esimerkiksi tiedostoja, kansioita ja ikkunoita. Valinta, raahaaminen ja avaaminen ovat esimerkkejä toiminnoista. (Gaver 1989, 77.) Jokainen kappale-toiminto -yhdistelmä muodostaa tapahtuman, jota on mahdollista kuvata vastaavilla äänitapahtumilla (Gaver 1989, 78). Vuorovaikutuksen lisäksi voidaan saadaan tietoa objektin ominaisuuksista. Ongelmana tässä on se, miten saada objektiivista tietoa objektin ominaisuuksista. Kuulija vertaa kuulemaansa muihin ääniin: esimerkiksi äänenkorkeudella ei ole absoluuttista arvoa, vaan se on aina käsitettävä suhteessa muiden äänten korkeuksiin. Eräs ratkaisu tähän voisi olla kaikkien käyttöliittymän objektien muodostama kokonaisäänimaisema, jonka sisällä voidaan vertailla eri ääniä toisiinsa eri konteksteissa. Keskeistä on se, voidaanko jokapäiväisiä ääniä liittää luonnollisesti ja merkityksellisesti käyttöliittymän tapahtumiin (Gaver 1989, 77). SonicFinder esittää pääasiassa

toisteista informaatiota (Gaver 1989, 82). Joissakin tapauksissa visuaalinen informaatio voidaan kuitenkin korvata kokonaan: napin painalluksesta voidaan antaa palautetta käyttäjälle pelkästään äänen avulla. Samalla tavalla palautetta voidaan antaa tehokkaasti äänen avulla muistakin toiminnoista (visuaalisen palautteen sijaan), jolloin käyttäjä voi keskittyä päätehtävään näkönsä avulla (Brewster, Wright, Dix & Edwards 1995, 48).

Gaver (1989, 89-90) ottaa ikonisten äänten kohdalla esiin kaksi ongelmaa: (a) minkälaisia ääniä pitäisi käyttää silloin kun kyseessä on vain tietokonemaailmassa esiintyvä tapahtuma, esimerkiksi virhe kirjoitettaessa levylle tai ikkunan avautuminen (ei avaudu kuten reaalimaailmassa, vaan zoomautuu isommaksi) ja (b) minkälaisia ääniä pitäisi käyttää sellaisissa käyttöliittymän tapahtumissa, joiden reaalimaailman vastineet eivät aiheuta ääntä tai aiheuttavat merkityksetöntä ääntä. Gaver esittää ratkaisuna ensin mainittuun ongelmaan elokuvallisten ääniefektien käyttöä. Elokuvien ääniefektit eivät liity tapahtumiin sattumanvaraisesti, vaan perustuvat kuuntelijoiden kykyyn yleistää tietämystään jokapäiväisistä äänitapahtumista. Näin ollen, vaikka ikkunat eivät SonicFinderissa avaudu, tämä tapahtuma muistuttaa muita jokapäiväisiä ääniä, esim. objektin äkillistä laajentumista. Myös Cohen (1993, 63-64) on ehdottanut lajityyppien käyttöä ("genre sounds"), jolloin käyttöliittymässä voitaisiin käyttää ääniä tutuista elokuvista tai televisio- ja radio-ohjelmista.

Jälkimmäiseen ongelmaan Gaver ehdottaa lähdemetaforien käyttöä. Esimerkiksi asiakirjaa kopioitaessa reaalimaailman kopiokone ei välitä informaatiota siitä, kuinka paljon kopioitavaa on vielä jäljellä. Tällainen ääni olisi kuitenkin tietokoneen käyttäjälle relevanttia tietoa. Tällöin on keksittävä jokin toinen äänitapahtuma kuvaamaan paremmin prosessin etenemistä, esimerkiksi veden kaatamisen ääni. Tällöin äänitapahtuma on ikonisessa suhteessa lähteeseensä, kun taas ääni-ikoni on metaforisessa suhteessa tapahtumaan. (Gaver 1989, 89-90.) Gaverin (1989, 91) mielestä näillä edellämainitulla kahdella tavalla voidaan

ikonisten äänten käyttörajoituksia hälventää. Ääniefektejä voidaan käyttää sellaisiin tapahtumiin, joita ei ole reaalimaailmassa, ja hyvinsuunnitellut efektit ovat silti kausaalisessa suhteessa tapahtumaan. Käyttämällä metaforisia ikonisia ääniä voidaan muodostaa analogioita äänitapahtumien välillä turvautumatta analogioihin tapahtumien ja äänen fysikaalisten attribuuttien välillä. (Gaver 1989, 91.)

Spatiaalisesta äänestä Gaver (1989, 92) toteaa, että ongelmana ei ole niinkään sijainti tilassa, vaan haluttujen objektien ja sijaintien löytäminen tilasta. Tällöin spatiaaliset ääni-ikonit voivat tarjota vihjeitä navigoinnille, tarjoamalla tietoa esimerkiksi äänellisen tilan koosta tai sen rajojen läheisyydestä. Kursori toimii tällöin tavallaan tutkana, joka tarkastelee ympäristöä. Tällainen lähestymistapa tuottaa sellaisia spatiaalisia vihjeitä mallimaailmasta, jotka eivät ole pelkästään visuaalisia. (Gaver 1989, 92.) Ikonisia ääniä kritisoidaan usein (ks. esimerkiksi Brewster 1994) niiden aiheuttamien rajoitusten vuoksi: tietyillä objekteilla tai tapahtumilla ei ole ikonista vastinetta. Miksei samalla tavalla kritisoida visuaalisia ikoneita, koska eiväthän nekään aina esitä sitä mitä niiden pitäisi? Ärsykkeeseen perustuva suora reaktio on nopeampi kuin symboliin koodatun informaation havaitsemiseen käytetty aika (Gibson 1966, 91). Tämä näkemys vahvistaa oletusta, että käyttöliittymässä tulisi käyttää mieluummin ikonisia kuin symbolisia ääniä, joita käsitellään seuraavaksi.

4.4 Symboliset äänet

Symboliset äänet (earcons) ovat abstrakteja, synteettisiä ääniä, joita voidaan käyttää strukturoidusti äänellisten viestien esittämiseen käyttöliittymässä (koskien objekteja tai niiden vuorovaikutusta) (Blattner ym. 1989, 13; Brewster ym. 1993, 222). Tutkijat käyttävät käsitettä vapaasti, viitaten sillä joskus sekä ikonisiin että symbolisiin ääniin. Koska "earconit" ovat useimmiten olleet musiikillisia ja synteettisiä ääniä, luokitellaan ne tässä symbolisiksi. Symboli ei

muistuta objektiaan (kuten ikoni), vaan merkitys syntyy siitä, mitä ihmiset ovat keskenään sopineet (Tarasti 1990, 30; Fiske 2000, 70). Tällöin äänen ja objektin välinen suhde on keinotekoinen. Gaverin (1989, 72) ja Herefordin ja Winnin (1994, 219) mukaan symboliset äänet eivät ole kovinkaan intuitiivisia, koska ne täytyy opetella ennen käyttöä; musiikillisten äänten liittäminen tietokoneen tapahtumiin on myös useimmiten satunnaista.

Symboliset äänet perustuvat motiiveihin, joita voidaan yhdistellä ja muunnella; niitä voidaan myös periyttää hierarkiassa alempana oleville äänitapahtumille (Blattner ym. 1989, 19; Beaudouin-Lafon & Gaver 1994, 50). Symbolisissa äänissä on Blattnerin ym. (1994, 453) mukaan se hyöty, että äänten ei tarvitse vastata kohdettaan. Tämä johtaa väistämättä siihen, että symboliset äänet täytyy opetella ennen käyttöä. Hälytysäänet ja signaalit ovat esimerkkejä symbolisista äänistä (Blattner ym. 1989, 14). Symbolien yhdistelmät - esimerkiksi objektin ja vuorovaikutuksen yhdistelmät - , saattavat olla helpommin toteutettavissa äänellä kuin visuaalisesti, koska ääni on seurausta vähintään kahden objektin vuorovaikutuksesta.

Schafer (1977, 170-178) luettelee esimerkkejä luonnossa ja elinympäristössämme esiintyvistä symbolista äänistä ja niihin mahdollisesti liitetyistä merkityksistä:

- vesi: puhdistava, uudistava, liikkeessä oleva, uudestisyntyvä, ikuinen, rytminen (esimerkiksi laineet)
- tuuli: ailahteleva, suuntaa vaihtava
- soittokellot: joko yhteenkerääminen (esimerkiksi kirkonkello) tai poisajaminen (pahojen henkien karkottaminen)
- torvet: voimakas, aggressiivinen, heijastavat viranomaisten arvovaltaa ja viittaavat voittoon
- sireenit: ahdistus ja hätä.

Tällaisten äänten hyödyntäminen käyttöliittymässä ei ole lainkaan keinotekoista, koska niiden merkitys on opittu jokapäiväisestä elämästä. Seuraavaksi käsitellään metaforia: perustuuhan suuri osa graafisesta käyttöliittymästä juuri niiden käytölle. Miten ääni voidaan liittää luonnolliseksi osaksi käyttöliittymän metaforista työpöytää?

4.5 Metaforiset äänet

Useimmat käyttäjät eivät tehtävää suorittaessaan ajattele tietokonetta koneena, vaan avaavat ikkunoita ja manipuloivat asiakirjoja (Gaver 1989, 85). Tällöin on kyse *metaforista*. Metaforassa käytetään hyväksi yhtäläisyyttä ja eroavuutta siirtämällä (assosioimalla) ominaisuuksia todellisuuden yhdeltä tasolta toiselle (Fiske 2000, 122, 127). Tällöin pyritään etsimään yhtäläisyyksiä eri tasojen välillä, mikä vaatii mielikuvitusta ja saattaa joskus synnyttää jopa surrealistisen vaikutelman (Fiske 2000, 130-131).

Laurelin (1991, 128) mukaan metaforien käyttö perustuu teoriaan, että kun käyttöliittymässä imitoidaan reaalimaailman objekteja, käyttäjät tietävät luonnostaan, miten toimia. Metaforien käyttö aiheuttaa Laurelin mielestä kuitenkin seuraavan ongelman: käyttöliittymän objekteilla ei ole joitakin reaalimaailman objektien fyysisiä ominaisuuksia (esimerkiksi painoa). Tämä johtaa siihen, että käyttöliittymäobjektit ovat reaalimaailman kaltaisia, mutta toisaalta puutteellisia, ja toisaalta täydennettyjä sellaisilla ominaisuuksilla, joita ei reaaliobjekteilla ole. Laurelin mielestä käyttöliittymämetaforat ovatkin *vertauksia*: vertaus kertoo käyttöliittymäobjektin olevan reaaliobjektin kaltainen, muttei aina sitä, millä tavalla nämä eroavat toisistaan. Metaforat siis toisaalta helpottavat käytön oppimista, mutta toisaalta sisältävät kognitiivisia epäloogisuuksia. (Laurel 1991, 128-130.)

Tärkeää äänen käytössä ovat yhdenmukaiset kausaalisuhteet äänen ja muiden modaliteettien välillä: äänen on vastattava suoritettua toimintoa (Laurel 1991, 160-161). Tämä viittaa siis jälleen audiovisuaaliseen sopimukseen ja väliaikaiseen jaksolliseen rakenteeseen.

Walker (1987) tutki kulttuurisen taustan, elinympäristön, iän ja musiikillisen taustan vaikutusta koehenkilöiden kykyyn valita sopiva visuaalinen metafora äänten akustisille parametreille. Tarkoituksena oli siis kuvata kuultua visuaalisen metaforan kautta kvalitatiivisesti. Koehenkilöt valitsivat seuraavat visuaaliset metaforat: taajuus - vertikaalinen sijainti; amplitudi - koko; aaltomuoto - kuvio; kesto - horisontaalinen pituus (Walker 1987, 496). Tulokset osoittivat, että musiikillinen koulutus vaikutti metaforien valitsemiseen enemmän kuin kulttuuri, ikä tai ympäristö (Walker 1987, 497-498). Tämä tutkimustulos vahvistaa näkemystä, että mikäli halutaan saavuttaa äänen käytöllä suurin mahdollinen käyttäjäryhmä, musiikillisten (symbolisten) äänten käyttö täytyy harkita tarkkaan. Äänten täytyy ainakin olla riittävän erilaisia.

Spatiaaliset metaforat ovat hyödyllisiä käyttöliittymässä, esimerkiksi jos informaatio järjestellään metaforisen rakennuksen huoneisiin (Wenzel 1994, 40). Useimmissa liikkuvan tietojenkäsittelyn sovelluksissa käytetään edelleen pöytätietokoneeseen liittyviä käsitteitä, kuten näppäimistöä, osoitinlaitteita ja graafisia elementtejä (Sawhney & Schmandt 1997, 109). Graafisissa käyttöliittymissä käytettyä työpöytä-metaforaa ei Brewsterin ym. (1998, 26) mielestä voida käyttää pienikokoisissa kannettavissa laitteissa näytön pienuuden takia. Myös työpöytä-metafora sisältää epäloogisuuksia, jotka on erikseen opeteltava: esimerkiksi reaalimaailman työpöytä ei sisällä ikkunoita (Hereford & Winn 1994, 214). Samoin, jos ääni muuttuu kovasta hiljaiseen tiedostoa tuhottaessa, se poikkeaa työpöytämetaforasta: asiat työpöydällä eivät tuosta vaan häivy hiljakseen näkyvistä (Gaver 1989, 87-88). Kartta saattaa olla hyvä spatiaalinen metafora, koska karttoja käytetään suuntautumiseen ja navigointiin

(Blattner ym. 1994, 458). Koska kartta on ahdettu täyteen visuaalista informaatiota, informaatiota saattaa jäädä piiloon. Jos taas sama informaatio sijoitellaan spatiaaliseen äänimaisemaan, informaation hukkuminen on epätodennäköisempää.

Navigoitaessa keskeistä on löytää haluttu objekti tilasta, jolloin graafisissa käyttöliittymissä usein käytetty puumainen hierarkia ei välttämättä ole kaikista käyttökelpoisin. Spatiaalisen äänen avulla on mahdollista saada aikaan abstrakti, mentaalinen tila - paikka - äänellä navigoinnille (yksityinen ääni-informaatioavaruus, kupla!). Tällöin voidaan vapautua näytön asettamista rajoituksista ja ahtaista raameista. Saue (2000, 1-4) esittää mielenkiintoisen metaforan: käyttäjä kävelee polkuja pitkin ja kuuntelee samalla lokaalisti ja globaalisti määriteltyjä äänitapahtumia. Informaatio on sijoiteltu tiettyihin sijainteihin äänimaisemaan, jossa käyttäjä voi kuunnella oman liikkumisensa aiheuttamia ääniä ("kävely"), lähiympäristön äänitapahtumia (paikalliset muuttujat), ympäristön tunnistamiseen tarvittavia ääniä (globaalit muuttujat) ja äänitapahtumien manipuloinnista aiheutuvia ääniä (esimerkiksi objektin valinta tai raahaaminen). Äänet ovat suhteessa kuuntelijan sijaintiin kullakin hetkellä, eli äänimaisema muuttuu jatkuvasti. Tämä seikka aiheuttaa kuitenkin ongelmia objektien sijainnin muistamisessa (ks. myös Kobayashi & Schmandt 1997). Saue esittää tähän kaksi ratkaisua: (a) kohdan merkitseminen visuaalisella merkillä ja (b) polun tekeminen kiinnostavan sijainnin läpi. Jälkimmäinen ratkaisu luo ajallisen jakson pisteitä, joka voi muodostaa mielekkään kokonaisuuden. Tällöin päästään jälleen kuuntelemaan Blyn (1982, 374) mainitsemaa merkityksellistä "laulua".

Sauen ajatukset saattavat pohjautua Howardin ja Ballaksen (1980) ja Warrenin ja Verbruggen (1984) artikkeleihin, joissa painotettiin väliaikaisen rakenteen ja syntaksin tärkeää osuutta merkityksen muodostamisessa. Sauen jako paikallisiin ja globaaleihin ääniin on myös mielenkiintoinen: käyttäjä voi kuunnella vain

niitä ääniä, jotka on paikallisesti määritelty tietyllä rajatulla alueella kokonaisäänimaisemassa. Tällainen jaottelu vihjaa jakoon tuttuihin ja tuntemattomiin ääniin. Sauen malli pohjautuu jokapäiväiseen kuuntelemiseen, huomion kiinnittymiseen paikallisiin tai globaaleihin ääniin, spatiaalisen äänen mahdollistamaan suuntautumiseen äänen suuntaan sekä äänitapahtuman ominaisuuksien havaitsemiseen (Saue 2000, 4).

Myös Schafer (1977, 212) puhuu "äänikävelystä" (soundwalk). Äänikävely käyttöliittymän kokonaisäänimaisemassa voi muodostaa metaforisen tilakokemuksen, jossa äänet merkitsevät ja kertovat käyttöliittymän tilasta ja toiminnoista. Schafer (1977, 211-212) puhuu "turistista" äänimaisemassa: äänet havaitaan paremmin vieraassa ääniympäristössä. Schaferin ideana on, että kuuntelija luokittelisi kuulemansa äänimaiseman äänet tietoisesti. Turistina oleminen on pelkkä välivaihe (ajattelutapa) havaita ympärillämme olevat äänet tehokkaammin.

Edellä käsiteltyjen ikonien, symbolien ja metaforien rajat eivät ole itsestäänselviä. Oleellista ei ehkä olekaan se, mihin kategoriaan ääni kuuluu. Äänet on paras hahmottaa jatkumona, jonka toisessa päässä ovat esittävät äänet ja toisessa päässä abstraktit äänet (Barger 1994, 59). Fiske (2000, 79) korostaa, että myös ikonien ymmärtäminen edellyttää jossain määrin niihin tottumista. Samaa mieltä on Tarasti (1990, 30) pohtiessaan, että myös ikonisuus on pohjimmiltaan käsitteellistä. Täysin puhdas ikoni lienee siten mahdottomuus. Tästä seuraa, että myös käyttöliittymän äänet täytyy opetella; täysin intuitiivinen käyttöliittymä saattaa olla käytännössä mahdoton toteuttaa. Seuraavaksi tarkastellaan äänitapahtumien erilaisia luokittelumahdollisuuksia.

4.6 Äänitapahtumien luokittelujärjestelmät

Jotta äänitapahtumia voitaisiin intuitiivisesti yhdistää käyttöliittymän toimintoihin, täytyy äänet ensin luokitella. Gaver (1993c, 9) analysoi äänitapahtumia kvalitatiivisesti ja jakaa ne kolmeen kategoriaan: (a) värähteleviin kappaleisiin (esimerkiksi askeleet), (b) nestemäisiin ääniin (esimerkiksi veden lorina tai kappaleen putoaminen veteen) ja (c) aerodynaamisiin tapahtumiin (esimerkiksi tuulen ääni). Gaverin (1993c, 20-21) tavoitteena on järjestellä jokapäiväisen kuuntelun äänitapahtumat yleiseen ja yksinkertaiseen, kuultavissa olevaan muotoon. Yksi tapa on järjestää ne kontekstin mukaan (esimerkiksi toimiston äänet). Gaver kuitenkin toteaa, ettei tämä tapa ole riittävä ja esittelee sen sijaan hierarkkisen viitehyksen. Koska äänitapahtuma syntyy materiaalien vuorovaikutuksen tuloksena (Gaver 1993c, 22), eräs mahdollisuus on sijoittaa materiaalit ylemmälle tasolle ja vuorovaikutuksen muodot alemmalle. Gaver (1993c, 26) kuitenkin toteaa, että esimerkiksi isku saattaa kertoa enemmän sen voimasta kuin materiaalista, tai enemmän materiaalista kuin vuorovaikutuksen muodosta.

Gaverin luokittelu on hyödyllinen, joskaan siinä ei tarkasti oteta kantaa siihen, mitä nämä materiaalit tai vuorovaikutukset merkitsevät käyttöliittymäkontekstissa (huomaa kuitenkin Gaverin "SonicFinder" (1989), jossa käytettiin menestyksellisesti mm. roskapöntön ääntä). Luokittelussa hylätään kontekstin osuus; elävässä elämässä tämä onkin perusteltua, koska myös toimistossa voi yllättäen kuulua mitä tahansa ääniä. Sen sijaan käyttöliittymässä kontekstin osuus äänten luokittelussa korostuu, koska virtuaalisen äänimaiseman ei täysin tarvitse noudattaa reaalitodellisuutta. Riittää, että tärkeimmät piirteet säilytetään.

Gibson (1966, 79-81) puolestaan jakaa äänimaiseman seuraavasti: (a) jatkuvat äänet (esimerkiksi vesiputous), (b) epäsäännölliset äänet (esimerkiksi tuuli) sekä

(c) äkilliset, odottamattomat äänet (esimerkiksi vieriminen ja hankaaminen). Tämän lisäksi Gibson mainitsee eläinten ja ihmisten tuottamat äänet (puhe, musiikki) sekä teknologisen aikakauden tuottamat koneelliset äänet. Kaikki nämä äänitapahtumat etenevät ajassa, ja useimmilla niistä on alku ja loppu. (Gibson 1966, 79-81.) Gibson siis jakaa äänet toisaalta niiden ilmenemistiheyden, toisaalta niiden aiheuttajien mukaan. Tällainen lähestymistapa saattaa olla käyttökelpoisempi kuin Gaverin hierarkkinen luokittelu. Gibsonin luokittelu lähenee pikemminkin Schaferin (1977) vastaavaa jaottelua: tällöin päästään vihdoin Schaferin edustamaan äänimaisemateoriaan.

Schaferin (1977, 137) mukaan hankaluutena äänten luokittelussa on se, ettei äänillä ole objektiivisia merkityksiä, vaan kullakin kuulijalla on kulttuuriset näkemyksensä niistä. Schafer (1977, 138-145) jakaa äänet semanttisesti seuraaviin luokkiin (ja nämä vielä alaluokkiin): (a) luonnolliset äänet (ilma, maa), (b) ihmisen tuottamat äänet (puhe, vaatteiden kahina), (c) kaupunki- ja maaseutuäänimaiseman äänet, (d) mekaaniset äänet (koneet, tuulettimet) ja (e) hiljaisuus. Tällainen lista on merkittävä vain mikäli sitä voidaan soveltaa käyttöliittymään. Sellaisenaan se on tähän tarkoitukseen liian laaja, koska siinä luokitellaan äänet vain niiden aiheuttajien mukaan. Schaferin mielestä (1977, 111) 1900-luvun musiikin huomiota herättävin piirre on se, että rajat musiikin ja ympäristön jokapäiväisten äänten välillä ovat hämärtyneet. Sovellettuna käyttöliittymään tämä merkitsee, että nykyajan ihminen on tottunut kaikenlaisten äänten yhdessäoloon. Tämä seikka vaikeuttaa käytettävien äänten valintaa entisestään.

On siis tarpeen miettiä, miten muuten äänet voitaisiin luokitella, jotta niitä voitaisiin soveltaa käyttöliittymään. Edellä äänitapahtumat on jaettu ainakin esittävyyden, abstraktisuuden tai assosiaation periaatteiden mukaan (ikonit, symbolit ja metaforat); materiaalien tai vuorovaikutuksen muodon mukaan; musiikillisten tai ei-musiikillisten äänten mukaan; luonnon tai mekaanisten

äänten mukaan. Nämä luokittelujärjestelmät eivät ole täysin vertailukelpoisia keskenään, koska ne sijoittuvat eri tasoille ja osin limittäin. Esimerkiksi luonnon äänet (siis ei-musiikilliset äänet) ovat useimmiten ikonisia, kun taas metaforiset äänet voivat mielikuvituksellisesti yhdistää piirteitä kaikista luokittelutasoista.

On kuitenkin vielä ainakin yksi tapa luokitella ääniä, jota ei käyttöliittymäkontekstissa ole aiemmin käsitelty. Schafer (1977) kehitti useita käsitteitä, joilla halusi luokitella äänimaisemassa kuuluvia ääniä. Onkin yllättävää, ettei Schaferin käsitteitä ole aiemmin yhdistetty käyttöliittymäkontekstiin. Käsitteet eroavat Schaferin edellä esitetystä jaottelusta: samalla kun ne kuuluvat jokapäiväiseen äänimaisemaamme, ne perustuvat äänten aiheuttajien lisäksi niiden rooleille äänimaisemassa ja ovat siten sovellettavissa spatiaaliseen käyttöliittymään. Lisäksi ne viittaavat tietyn yhteisön olemassaoloon ja tapahtuvat siten aina tietyssä kontekstissa. Äänimaisemaa voidaan Schaferin (1977, 9-10, 43-44) mukaan kuvata seuraavilla käsitteillä:

- *Perusääni* on ääni, joka kuuluu yhteisössä jatkuvasti tai tarpeeksi usein muodostaen taustan muille äänille.
- *Äänellinen maamerkki* on yhteisössä ilmenevä ainutlaatuinen ja merkittävä ääni, jota on syytä suojella. Määritelmä viittaa myös tietyn tilan tunnistamiseen äänen avulla.
- *Signaali* tarkoittaa ääntä, johon kiinnitetään erityisesti huomiota. Esimerkkinä tästä ovat varoitussignaalit.
- *Hi-fi* viittaa ympäristöön, jossa äänet ovat kuultavissa sekoittumatta tai peittymättä toistensa alle.
- *Lo-fi* viittaa ympäristöön, jossa äänet sekoittuvat tai peittyvät toistensa alle. Lo-fi -ympäristö sai alkunsa teollisen vallankumouksen seurauksena (1700-luvun lopulta lähtien) ja vahvistui entisestään sähköisen vallankumouksen aikana (Schafer 1977, 71). Hiljainen hi-fi -maailma on siis muuttunut kovaääniseksi lo-fi -maailmaksi useiden vuosisatojen aikana.

Käsitteet tarjoavat luontevan ympäristön spatiaalisen äänen käytölle. Schafer puhuu yhteisöstä; tässä yhteydessä se voidaan rinnastaa käyttöliittymän muodostamaan akustiseen tilaan. Schaferin (1977, 214) mukaan *akustinen tila* tarkoittaa sitä aluetta, jonka sisällä äänet kuuluvat: esimerkiksi ihmisen tuottama akustinen avaruus rajoittuu siihen alueeseen, jossa hänen huutonsa kuuluu. Koneiden avulla tätä akustista aluetta voidaan suurentaa ja myös täyttää ahtaammin. (Schafer 1977, 214.) Akustisen tilan äänet merkitsevät jotain vain niissä (virtuaalisissa) ympäristöissä ja konteksteissa, jotka on sidottu käyttöliittymän tapahtumiin. Tällöin siis voidaan yhdistää äänen funktio sen sijaintiin.

Luonnon äänimaisemassa jokaiselle äänitapahtumalle on aikansa; Schafer väittää, että tällainen vuoropuhelu on katoamassa urbaanista maailmasta, jolloin päädytään lo-fi -äänimaisemaan ja äänten kakofoniaan (Schafer 1977, 229, 237). Schaferin mukaan aiemmissa yhteisöissä suurin osa äänistä oli yksittäisiä ja niillä oli selkeä alku ja loppu, kun taas nykyaikana suurin osa äänistä on jatkuvia (Schafer 1977, 78). Schafer (1977, 88-90) toteaa myös, että aiemmin kaikki äänet olivat alkuperäisiä, eli ne tapahtuivat tiettynä aikana tietyssä paikassa vain kerran. Sähköisenä aikakautena on mahdollista erottaa ääni äänilähteestään ja siten sen alkuperäisestä kontekstistaan. Tällöin on mahdollista simuloida mitä tahansa ääniympäristöä ja tehdä siitä kannettava, mukana kulkeva kokonaisuus (Schafer 1977, 91).

Schaferin ajatuksista saa dramaattisen vaikutelman, että ihmiset pakenevat ulkomaailman modernia lo-fi -ääniympäristöä koteihinsa löytääkseen sieltä rauhallisemman hi-fi -äänimaiseman; samalla tavalla käyttöliittymässä voitaisiin järjestää ihanteellinen keinotodellisuus (yksityinen kupla), jossa kaikki äänet kuuluvat ja ovat merkityksiltään yksiselitteisiä, koska ne on otettu päivittäisestä ympäristöstä ja vain järjestetty paremmin. Koska reaalitodellisuudessa

äänimaisemaa saattaa olla mahdotonta muuttaa kovinkaan paljon paremmaksi, käyttöliittymän virtuaalitodellisuus voi tarjota tähän toisen mahdollisuuden. Schafer (1977, 217-219) pohtii myös, että äänet sisätilassa saattavat merkitä yksityisyyttä, kun taas kaikuisat äänet saattavat merkitä jatkuvuutta tai auktoriteettia. Juuri tällaiset merkitykset ovat tärkeitä - ja toistaiseksi vähemmän käytettyjä - käyttöliittymässä. Ne myös viittaavat suoraan spatiaaliseen ääneen. Näiden löydösten perusteella päädytään seuraavassa luvussa kokoamaan yhteen tutkimuksen tulokset.

5 KESKUSTELUA

Luvussa esitellään keskeisimmät tulokset, joita sitten tarkastellaan alussa asetettujen tutkimusongelmien valossa ja aiempiin tutkimuksiin suhteuttaen.

5.1 Tulokset

Tutkimus on perustunut olettamukseen kuulohavaintojärjestelmän kaksijakoisesta funktiosta: aktiivisesta äänen suunnan havaitsemisesta sekä äänilähteen tunnistamisesta. Tällä perusteella äänen suunnan tuottama merkityssisältö on pyritty yhdistämään käyttöliittymän toimintoihin ja objekteihin piilossa olevan informaation paljastamiseksi. Vaikka sijainti ei olekaan ainoa tekijä äänten erottelun ja ryhmittelyn kannalta, se on valittu käsiteltäväksi juuri sen tarjoaman merkityssisällön vuoksi. Äänivirta mentaalisena kokemuksena voidaan rinnastaa visuaalisen objektin kokemiseen, koska yksittäisestä äänilähteestä peräisin olevat äänet tulevat yleensä yhdestä sijainnista äänellisestä avaruudesta. Muistin aiheuttamien rajoitusten vuoksi käsiteltäväksi on rajattu paikallaan pysyvä ääni.

Kuulo- ja näköaistin vertailu paljasti niiden perimmäiset - joskin tässä yksinkertaistetut - erot, samoin kuin kummankin vahvimmat puolet. Koska kuuloaistimme avulla saamme tietoa taustalla tapahtuvista prosesseista, voimme päättää, mitkä osat ympäristöstämme tuodaan näköaistin alueelle. Tästä seuraa, että spatiaalinen ääni auttaa löytämään näytöltä visuaalisen objektin nopeammin, kun kohde on näkökentän ulkopuolella tai muuten piilossa esimerkiksi toisen ikkunan takana (tai ei mahdu pienikokoiselle näytölle). Suuntautumisen lisäksi olennainen ero näkö- ja kuuloaistin välillä on ajallisuus: ääni perustuu vaihteluihin ajassa ja viittaa siten yleensä tilan tarkkailuun tai muutokseen.

Kuulo- ja näköaistin eroista johtuen - ja siten toisiaan täydentäen - kokemus syntyy äänellisen ja visuaalisen informaation yhteistyönä.

Synteettisen spatialisoinnin toteutuksessa olennaisimmiksi ongelmiksi osoittautuivat peilikuvaäänet, paikantamisvirheet ja tietokoneen suorituskyvyn ylittävä datamäärä (vimeksimainitun ongelman osalta tilanne tosin edistyy koko ajan tekniikan myötä). Vaihe- ja voimakkuuserot korvien välillä, korvalehden ja -käytävän vaikutus sekä pään aiheuttama äänen vaimeneminen ovat tärkeimmät tekijät, jotka muuntavat äänen spektriä muodostaen siirtofunktion, jonka perusteella ääni paikannetaan. Paikantamisvirheet näyttävät keskittyvän siirtofunktion ympärille. Näitä paikantamisvirheitä voidaan tosin huomattavasti vähentää, jos äänen toistossa käytetään kuulokkeita.

Paikantaminen on monimutkainen prosessi, jonka kattava selvittäminen tässä yhteydessä ei ole ollut mahdollista. Tästä syystä tutkimuksen psykoakustista osuutta on väistämättä jouduttu yksinkertaistamaan kuitenkin siten, että keskeisimmät seikat on käsitelty. Toteutukseen liittyvät ongelmat ovat kuitenkin voitettavissa; tästä ovat osoituksena lukuisat spatiaalista ääntä hyödyntävät sovellukset (lähinnä kuitenkin laboratorio-olosuhteissa, ei kaupallisessa käytössä).

Spatiaalisia ääniä käytettäessä on erityisen tärkeää, että erilliset peräkkäiset äänitapahtumat noudattavat väliaikaista jaksollista rakennetta. Peräkkäisten äänitapahtumien on kielen tavoin perustuttava syntaksiin; syntaksi muodostuu niistä mahdollisista väliaikaisista suhteista, joista jakso äänitapahtumia voi keskenään muodostaa järkevän kokonaisuuden. Järkevä kokonaisuus syntyy reaalitodellisuudessa äänilähteiden sanelemassa järjestyksessä ja kestää äänilähteiden ominaisuuksista riippuvan ajan. Tätä seikkaa voidaan hyödyntää jakamalla käyttöliittymän toiminnot äänitapahtumien jaksoiksi: jakso koostuu toiminnoista, jotka käyttäjä tekee suorittaakseen jonkin tehtävän.

Ensimmäisen tutkimusongelman mukaisesti pyrittiin selvittämään, miten spatiaalista ääntä voitaisiin käyttää grafiikan kanssa tehtävän suorituksessa ja piilossa olevien objektien tai tapahtumien esittämisessä. Tulosten mukaan spatiaalinen ääni tarjoaa uuden ulottuvuuden niihin funktioihin, joissa ääni voi käyttöliittymässä toimia. Kun äänen tarkoitus on esimerkiksi antaa palautetta käyttäjän toiminnasta, miksi tyytyä monauraaliseen tai stereoääneen? Spatiaalisen äänen käyttöä puolustavat seuraavat seikat:

- Toteutus on teknisesti mahdollinen.
- Jokapäiväinen äänimaisemamme on spatiaalinen.
- Spatiaalinen ääni osoittaa, minne katsoa.
- Spatiaalinen ääni voi tarjota informaatiota akustisen tilan ominaisuuksista ja objektien sijainnista.
- Informaatiota voidaan järjestellä 360 asteen alueelle, jolloin visuaalisen informaation määrä vähenee.

Lisäksi, jokapäiväiset spatiaaliset äänet voivat välittää hienovaraisempaa informaatiota äänitapahtumasta tai vuorovaikutuksen muodosta kuin perinteisten musiikillisten instrumenttien tuottamat äänet (joita käyttöliittymässä on perinteisesti käytetty monauraalisesti tai stereona). Jäljelle jää siten vain sopivan luokittelujärjestelmän valitseminen: tällöin voidaan saada selville, minkälaisia ääniä kannattaisi käyttää missäkin käyttöliittymän tapahtumassa. Tämä on kaikkea muuta kuin helppo tehtävä.

Tutkimuksessa on tarkasteltu useita erilaisia - sekä hierarkkisia että kontekstipohjaisia - äänitapahtumien luokittelujärjestelmiä. Niiden vastakkainasettelun perusteella käyttökelpoisimmaksi osoittautui Schaferin (1977) esittämä luokittelu: äänitapahtumien jako perusääniin, äänellisiin maamerkkeihin ja signaaleihin viittaa tietyssä kontekstissa olemassaolevaan akustiseen tilaan. Tällainen jaottelu tarjoaa siten yhden mahdollisen vastauksen

tutkimusongelmaan, pitäen sisällään sekä äänen roolin - merkityksen - että spatiaalisuuden käyttöliittymässä. Näköaistin osuus tässä on se, että käyttäjä voi kääntää katseensa objektiin (tai reagoida muulla tavoin) ainoastaan silloin, kun siihen todella on tarvetta. Näin näköaisti vapautuu sen hetkiseen tärkeään tehtävään, eikä kuormitu turhalla informaatiolla. Spatiaalinen ääni voi ilmaista tämän tarpeen äänen sijainnin ja merkityksen - esimerkiksi kiireellisyyden (signaali) tai äänellisen maamerkin - avulla; merkitykset kun on opittu jokapäiväisestä elämästä. Tällöin voidaan saavuttaa intuitiivinen kuvallisen ja äänellisen informaation yhdistelmä; seikka jota on hyödynnetty peleissä jo vuosikausia. Spatiaalisen äänen muodostamassa "kannettavassa kuplassa" informaatiota voidaan järjestellä laajemmalle alueelle kuin visuaalista informaatiota.

On kuitenkin otettava huomioon, että moderni jokapäiväinen äänimaisema voi sisältää myös musiikillisia ääniä: ero on äänten kokemisessa. Oleellista on se, että äänet ja äänilähteet ovat mahdollisimman ikonisessa suhteessa toisiinsa, jolloin niiden opettelemiseen kuluu vähemmän aikaa. Oli kyseessä sitten musiikilliset tai jokapäiväiset äänet, niiden avulla voidaan esimerkiksi tunnistaa jokin metaforinen akustinen tila - kyseessä on tällöin äänellinen maamerkki. Kontekstin merkitys objektien tunnistamisessa ja merkityksenannossa korostuu etenkin silloin, kun eri fysikaaliset tapahtumat aiheuttavat samankaltaisia ääniä. Aina ei kuitenkaan ole välttämätöntä objektiivisesti tunnistaa äänilähdettä: riittää, että se on reaaliobjektin kaltainen. Käyttöliittymän virtuaalitodellisuudessa voidaan korostaa ja liioitella sellaisia piirteitä, jotka tekevät ääniobjektin merkityksestä yksiselitteisen käyttöliittymäkontekstissa.

Seuraavaksi esitellään toisen tutkimusongelman mahdollista ratkaisua. Tavoitteena oli selvittää, voidaanko käyttöliittymän ulkopuolisia spatiaalisia ääniä hyödyntää piiloinformaation esittämisessä. "Ulkopuolisilla" äänillä tarkoitettiin sellaisia ääniä, jotta eivät suoraan liity käyttöliittymän toimintoihin.

Yhtenä ratkaisuna voidaan pitää elokuvallisten ääniefektien käyttöä. Tälle voidaan esittää kaksi perustelua: ensinnäkin, ihmisten voidaan olettaa kykenevän yleistää tietämystään jokapäiväisistä äänitapahtumista. Tällöin riittää se, että äänitapahtumat muistuttavat jokapäiväisiä ääniä: kyse on siten metaforisesta assosiaation periaatteesta. Toiseksi: voidaan olettaa, että tietyt äänet ovat tuttuja elokuvista tai televisio- ja radio-ohjelmista. Tällöin käyttöliittymässäkin voitaisiin päästä lähemmäksi tietokonepelien maailmaa – sitä maailmaa, jossa elokuvat ovat oleskelleet jo useiden kymmenien vuosien ajan. Tästä kuitenkin seuraa, että käytön opettelemiselle täytyy varata hieman enemmän aikaa.

Edellä mainittu "ulkopuolinen" ääni voi viitata myös toiseen merkitykseen: nimittäin näytön ulkopuolella, näkymättömissä olevan objektin aiheuttamaan ääneen. Tällä on kaksi seurausta: näytön ulkopuolinen ääni viittaa (a) piilossa olevan informaation olemassaoloon sekä (b) audiovisuaaliseen sopimukseen, koska sulkiessamme silmämme - tai katsoessamme muualle - näytön ulkopuoliset äänet muuttuvat näytöllä oleviksi ääniksi. Tämä on yksi osoitus siitä, että elokuvallisella äänellä ja siihen liittyvillä teorioilla on paljon tarjottavana myös käyttöliittymälle. Seuraavassa kohdassa analysoidaan tuloksia tarkemmin suhteessa aiempiin tutkimuksiin.

5.2 Johtopäätökset

Tutkimus osoittaa, että käyttöliittymässä on mahdollista muodostaa metaforinen akustinen tila, jossa käyttäjä kuuntelee ikonisia spatiaalisia ääniä ja reagoi niihin niiden sisältämän informaatiosisällön mukaisesti. Spatiaalisten jokapäiväisten äänten avulla voidaan paljastaa informaatiota, joka muutoin jäisi visuaalisesti piiloon. Tutkimuksen kontribuutiona voidaan pitää sitä, että siinä on koottu erillisiä tutkimustuloksia pyrkien yhdistämään sitä tietämystä, joka muutoin jäisi insinöörien, musiikkitieteen, kognitiotieteen tai viestintätieteen tutkijoiden erityisalaksi. Aiemmassa tutkimuskirjallisuudessa on ollut silmiinpistävää se, että

niissä on keskitytty toisaalta joko symbolisten äänten käyttäjätestaukseen ilman taustalla olevaa teoriaa tai teoreettiseen pohdiskeluun äänen roolista visuaalisen kuorman keventämisessä ilman empiiristä testausta. Asioita ei ole kyetty luontevasti yhdistämään.

Tässä tutkimuksessa ei ole muodostettu varsinaista omaa uutta teoriaa tai mallia; pikemminkin kyse on uudesta näkökulmasta ja asioiden luovasta yhdistämisestä. Olennaisimpana kontribuutiona voidaan pitää akustisen ekologian ja sen sisältämien spatiaalisten äänimaisemakäsitteiden - samoin kuin elokuvallisten ääniefektien - yhdistämistä käyttöliittymäkontekstiin. Tällaiset kytkennät ovat erityisen hyödyllisiä, koska ne ovat sovellettavissa laajalti eri järjestelmiin ja laitteisiin. Tekniikka tätä varten on ollut valmiina jo vuosia. TAULUKOSSA 2 on vielä koottu yhteen tutkimuksen tärkeimmät tulokset.

Toisaalta tutkimuksen varjopuolena taas on aiheen laajuus ja abstrakti luonne, jolloin on ollut vaikea välttää liiallista yksinkertaistamista ja yleistämistä. Tulokset kuitenkin osoittavat, että ääniä todella voidaan käyttää laajan käyttäjäryhmän hyväksi; tärkeintä on, ettei musiikillisen koulutuksen puute muodostu erottavaksi tekijäksi. Mahdollisimman laaja käyttäjäryhmä tulisi tietenkin olla tavoitteena, jolloin on perusteltua käyttää sellaisia jokapäiväisiä ääniä, joihin useimmat ovat tottuneet. Tutkimuksessa ei ole esitetty kovinkaan konkreettisia esimerkkejä (muutamia sovelluksia lukuunottamatta): tämä on perusteltua siksi, että alalta puuttuu kokonaisuuden kattava teoria. Juuri tästä syystä tutkimusmenetelmäksi on valittu käsitteellinen synteesi. Aihetta olisi voitu käsitellä myös kaupallisen nykytekniikan näkökulmasta, jolloin saavutettu hyöty olisi ehkä ollut selvemmin osoitettavissa. Näin ei kuitenkaan haluttu tehdä: inhimillisen näkökulman merkitystä ei voi liikaa painottaa. Tekniikan on aina taivuttava ihmisen hyväksi, eikä se saa rajoittaa toteutusta.

TAULUKKO 2. Tutkimuksen tulokset

ONGELMA	TULOKSET
1. Spatiaalinen ääni piiloinformaation välittäjänä	- Äänitapahtumien on noudatettava väliaikaista jaksollista rakennetta. - Toteutus on teknisesti mahdollinen. - Spatiaalinen ääni osoittaa, minne katsoa. - Informaatiota voidaan järjestellä 360 asteen alueelle, jolloin visuaalisen informaation määrä vähenee. - Käyttöliittymän äänet voidaan luokitella spatiaalisiin perusääniin, äänellisiin maamerkkeihin ja signaaleihin, jolloin saadaan vihjeitä tietyn akustisen tilan ominaisuuksista ja objektien sijainneista tietyssä kontekstissa.
2. Käyttöliittymän ulkopuolinen spatiaalinen ääni	- Elokuvallisten ääniefektien käyttö.

Vaikka tässä tutkimuksessa esitetyt tulokset saattavat kirjoitushetkellä olla utopiaa kaupallisissa sovelluksissa, se ei kuitenkaan saa olla tekosyy äänen täydelliselle hylkäämiselle käyttöliittymistä. Tutkimuksen päätteeksi yhteenvedossa käydään lyhyesti läpi tutkimuksen rakenne ja tulokset sekä pohditaan jatkotutkimusaiheita.

6 YHTEENVETO

Tutkimuksessa on tarkasteltu spatiaalista ääntä graafisen käyttöliittymän laajentajana. Ääntä ei ole toistaiseksi hyödynnetty käyttöliittymässä riittävästi. Tarkoituksena on ollut selvittää, (a) miten spatiaalista ääntä voitaisiin käyttää tehokkaasti grafiikan kanssa piiloinformaation välittämisessä ja tehtävän suorituksessa ja (b) miten sellaista ääntä, jolla ei ole selvää vastinetta käyttöliittymässä, voitaisiin käyttää piilossa olevan informaation esittämisessä. Käsitteellis-teoreettisen analyysin avulla on pyritty sijoittamaan keskeiset käsitteet laajempaan viitekehykseen. Tutkimukseen on sisällytetty myös äänten luokittelujärjestelmien vertailu.

Tutkimus on jakaantunut kahteen pääteemaan: (a) spatiaalisen äänen tuottamaan merkityssisältöön ja (b) ympäristön spatiaalisten äänten hyödyntämiseen ja liittämiseen käyttöliittymän objekteihin ja toimintoihin. Taustaolettamuksena on ollut, että kuuloaistin funktiona on äänen suunnan havaitseminen ja äänilähteen tunnistaminen. Tutkimuksessa on vertailtu ensin kuulo- ja näköaistia, määritelty äänen paikantamiseen vaikuttavat seikat, esitetty sitten synteettisen spatialisoinnin yleisperiaate ja toteutukseen liittyvät ongelmat. Keskeisimpiä ongelmia toteutuksessa ovat peilikuvaäänet, paikantamisvirheet ja tietokoneen suorituskyvyn ylittävä datamäärä. Paikantamisvirheet aiheutuvat epätarkasti mitatusta siirtofunktiosta. Psykoakustisen tarkastelun jälkeen ääntä on tarkasteltu käyttöliittymäkontekstissa semioottisessa viitekehyksessä.

Tulokset osoittavat, että spatiaalinen ääni voi tuoda uuden ulottuvuuden niihin funktioihin, joissa ääni voi käyttöliittymässä toimia. Tällöin spatiaaliset ääni-ikonit voivat tarjota vihjeitä navigoinnille, tarjoamalla tietoa esimerkiksi akustisen tilan koosta tai äänellisestä maamerkistä, joka viittaa tiettyyn

kontekstiin. Sen lisäksi, että spatiaalinen ääni osoittaa minne katsoa, se voi välittää hienovaraista informaatiota prosessin kulusta, akustisen tilan ominaisuuksista tai tapahtuman kiireellisyydestä äänen sijainnin ja kategorian avulla. Tärkeää äänen käytössä ovat audiovisuaalisen sopimuksen ja väliaikaisen jaksollisen rakenteen huomioon ottaminen. Tutkimuksen perusteella on selvää, että mikäli ääniä halutaan menestyksellisesti liittää käyttöliittymän toimintoihin, on otettava huomioon jokapäiväinen kuunteleminen, spatiaalisen äänen mahdollistama suuntautuminen äänen suuntaan sekä äänitapahtuman ominaisuuksien havaitseminen.

Teoreettisen tarkastelun jälkeen on otettava konkreettisempi näkökulma. Jatkossa on tärkeää empiirisesti tutkia sitä, miten tämän tutkimuksen perusteella spatiaalisia ympäristön ääniä voitaisiin - käytännön tasolla - kytkeä käyttöliittymän tapahtumiin. Jatkotutkimusaiheita on runsaasti psykoakustiikan, käytännön sovellusten ja akustisen ekologian piirissä. Näistä akustinen ekologia tarjonnee suurimmat haasteet ja hyödyt, koska akustisen ympäristömme äänten käyttäminen käyttöliittymässä voi johtaa intuitiiviseen lopputulokseen.

Erityisesti liikkuvan tietojenkäsittelyn yleistymisen ja laitteiden näyttöjen pienentymisen seurauksena osa visuaalisesta informaatiosta on korvattava äänen avulla. Tällöin spatiaalinen ääni tarjoaa 360 asteen ulottuvuuden informaation esittämiseen. Äänen käyttö käyttöliittymissä tulee melko varmasti lisääntymään tulevaisuudessa huimasti. Toiveena tietenkin on, että spatiaalisen äänimaailman kehitys ja hyväksikäyttö voi murtaa näköaistin ylivallan ja saattaa virtuaalitodellisuuden lähemmäksi todellisuutta - audiovisuaalisen sopimuksen mukaisesti.

LÄHTEET

Ackerman, M. S., Starr, B., Hindus, D. & Mainwaring, S. D. 1997. Hanging on the `wire: a field study of an audio-only media space. ACM Transactions on Computer-Human Interaction, 4(1), 39-66.

Albers, M. C., Bergman, E. 1995. The audible web: auditory enhancements for Mosaic. Teoksessa Conference companion on Human factors in computing systems, Denver, CO, USA, May 7-11, 1995, New York: ACM Press, 318-319.

Alten, S. R. 1999. Audio in media. 5. painos. USA: Wadsworth publishing company.

Anderson, J. R. 1980/2000. Cognitive psychology and its implications. 5. painos. New York: Worth Publishers.

Arons, B. 1992. A Review of the cocktail party effect. Journal of the American Voice I/O Society, 12 (July 1992), 35-50.

Ballas, J. A. 1993. Common factors in the identification of an assortment of brief everyday sounds. Journal of Experimental Psychology: Human Perception and Performance, 19(2), 250-267.

Ballas, J. A. 1994. Delivery of information through sound. Teoksessa G. Kramer (toim.) Auditory Display: Sonification, Audification and Auditory Interfaces, Reading, MA, USA: Addison-Wesley Publishing Company, 79-94.

Ballas, J. A. & Howard, J. A., Jr. 1987. Interpreting the language of environmental sounds. Environment and Behavior, 19(1), 91-114.

Ballas, J. A. & Mullins, R. T. 1991. Effects of context on the identification of everyday sounds. Human Performance, 4(3), 199-219.

Barger, R. 1994. Pattern and reference in auditory display. Teoksessa G. Kramer (toim.) Auditory Display: Sonification, Audification and Auditory interfaces, Reading, MA, USA: Addison-Wesley Publishing Company, 151-165.

Beaudouin-Lafon, M., Gaver, W. 1994. ENO: Synthesizing structured sound spaces. Teoksessa Proceedings of the Seventh Annual Symposium on User Interface Software and Technology, Marina del Rey, CA, USA, Nov. 2-4, 1994, New York: ACM Press, 49-57.

Begault, D. R. 1991. Challenges to the successful implementation of 3-D sound. Journal of the Audio Engineering Society, 39(11), 864-870.

Begault, D. R. 1994. 3-D sound for virtual reality and multimedia. Cambridge MA: Academic Press.

Begault, D. R. 1999. Auditory and non-auditory factors that potentially influence virtual acoustic imagery. Teoksessa Proceedings of the Audio Engineering Society 16th international conference on spatial sound reproduction, Rovaniemi, Finland, April 10-12, 1999, 13-26.

Blattner, M. M., Greenberg, R. M. & Kamegai, M. 1992. Listening to turbulence: an example of scientific audiolization. Teoksessa M. Blattner & R. Dannenberg (toim.) Multimedia Interface Design, New York: ACM Press, 87-102.

Blattner, M. M., Papp III, A. L. & Glinert, E. P. 1994. Sonic enhancements of two-dimensional graphic displays. Teoksessa G. Kramer (toim.) Auditory Display: Sonification, Audification, and Auditory Interfaces, Reading, MA, USA: Addison Wesley Publishing Company, 447-470.

Blattner, M. M. & Sumikawa, D. A. & Greenberg, R. M. 1989. Earcons and icons: Their structure and common design principles. Human-Computer Interaction 4(1), 11-44.

Bly, S. 1982. Presenting information in sound. Teoksessa Proceedings on Human Factors in Computer Systems, Gaithersburg, Maryland, USA, 1982, New York: ACM Press, 371-375.

Bregman, A. S. & Campbell, J. 1971. Primary auditory stream segregation and perception of order in rapid sequences of tones. Journal of Experimental Psychology, 89(2), 244-249.

Brewster, S. A. 1991. Providing a model for the use of sound in user interfaces. (Tech. Rep. No. YCS169). University of York, Department of Computer Science, 1-54.

Brewster, S. A. 1994. Providing a structured method for integrating non-speech audio into human-computer interfaces. University of York, UK. Väitöskirja.

Brewster, S. A., Wright, P. C. & Edwards, A. D. N. 1993. An evaluation of earcons for use in auditory human-computer interfaces. Teoksessa Proceedings of the Conference on Human Factors in Computing Systems, Amsterdam, The Netherlands, April 1993, Boston: Addison-Wesley Longman Publishing Company, 222-227.

Brewster, S. A., Wright, P. C., Dix, A. J. & Edwards, A. D. N. 1995. The sonic enhancement of graphical buttons. Teoksessa K. Nordby, P. Helmersen, D. Gilmore, & S. Arnesen (toim.) Proceedings of the IFIP International Conference on Human-Computer Interaction, Lillehammer, Norway, June 25-29, 1995, London: Chapman & Hall, 43-48.

Brewster, S. A., Wright, P. C & Edwards, A. D. N. 1995. The application of a method for integrating non-speech audio into human-computer interfaces. (Tech. Rep. No. YCS253). University of York, Department of Computer Science, 1-19.

Brewster, S.A., Leplatre, G. & Crease, M.G. 1998. Using non-speech sounds in mobile computing devices. Teoksessa C. Johnson (toim.) Proceedings of the First Workshop on Human Computer Interaction with Mobile Devices, Glasgow, UK, Department of Computing Science, University of Glasgow, May 21-23, 1998, 26-29.

Broadbent, D. E. 1958. Perception and communication. New York: Pergamon.

Brown, M. L., Newsome, S. L. & Glinert, E. P. 1989. An experiment into the use of auditory cues to reduce visual workload. Teoksessa Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Austin, TX, USA, April 30-May 4, 1989, New York: ACM Press, 339-346.

Burgess, D. A. 1992a. Real-time audio spatialization with inexpensive hardware. Teoksessa Proceedings of the Third International Conference on Signal Processing Applications and Technology, Boston, MA, USA, Nov 2-5, 1992.

Burgess, D. A. 1992b. Techniques for low cost spatial audio. Teoksessa Proceedings of the fifth annual ACM symposium on user interface software and technology, Monteray, CA, USA, Nov. 15-18, 1992, New York: ACM Press, 53-59.

Buxton, W. 1989. Introduction to this special issue on nonspeech audio. Human-Computer Interaction, 4(1), 1-9.

Buxton, W. 1990. Using our ears: an introduction to the use of nonspeech audio cues. Teoksessa E. J. Farrell (toim.) Extracting Meaning from Complex Data: Processing, Display, Interaction, Vol. 1259, 124-127.

Carter, R. C. 1982. Visual search with color. Journal of Experimental Psychology: Human Perception and Performance, 8(1), 127-136.

Chion, Michel. 1990. Audio-vision: sound on screen. New York: Columbia University Press.

Cohen, J. 1993. "Kirk here": Using genre sounds to monitor background activity. Teoksessa S. Ashlund, K. Mullet, A. Henderson, E. Hollnagel, & T. White (toim.) Proceedings of the Conference on Human Factors in Computing Systems, Amsterdam, The Netherlands, Apr. 24-29, 1993, New York: ACM Press, 63-64.

Cohen, J. 1994. Monitoring background activities. Teoksessa G. Kramer (toim.) Auditory Display: Sonification, Audification and Auditory Interfaces, Reading, MA, USA: Addison-Wesley Publishing Company, 499-531.

Dannenberg, R. & Blattner, M. 1992. Introduction: the trend toward multimedia interfaces. Teoksessa M. Blattner & R. Dannenberg (toim.) Multimedia Interface Design, New York: ACM Press, xvii-xxv.

Deutsch, D. 1980. The processing of structured and unstructured tonal sequences. Perception and psychophysics, 28(5), 381-389.

Edwards, A. D. N. 1988. The design of auditory interfaces for visually disabled users. Teoksessa Proceedings of the Conference on Human Factors and Computing Systems, Washington D.C., USA, May 1988, New York: ACM Press, 83-88.

Fernström, M. & Bannon, L. Multimedia browsing. 1997. Position paper for the CHI'97, Workshop on Navigation on Electronic Worlds, Atlanta, GA, USA, March 23-24, 1997, 1-10.

Fiske, J. 2000. Merkkien kieli: johdatus viestinnän tutkimiseen. Suomeksi toimittaneet V. Pietilä, R. Suikkanen & T. Uusitupa. 6. painos. Jyväskylä: Gummerus Kirjapaino Oy.

Fitch, W. T. & Kramer, G. 1994. Sonifying the body electric: superiority of an auditory over a visual display in a complex, multivariate system. Teoksessa G. Kramer (toim.) Auditory Display: Sonification, Audification, and Auditory Interfaces, Reading, MA, USA: Addison-Wesley Publishing Company, 307-326.

Gaver, W.W. 1986. Auditory icons: Using sound in computer interfaces. Human-Computer Interaction, 2(1), 167-177.

Gaver, W.W. 1989. The Sonicfinder: an interface that uses auditory icons. Human-Computer Interaction, 4(1), 67-94.

Gaver, W.W. 1993a. How do we hear in the world?: explorations in ecological acoustics. Ecological Psychology 5(4), 285-313.

Gaver, W. W. 1993b. Synthesizing auditory icons. Teoksessa Proceedings of INTERCHI'93 Conference on Human Factors in Computing Systems, Amsterdam, The Netherlands, Apr. 24-29, 1993, Reading, MA, USA: ACM Press/Addison-Wesley, 24-29.

Gaver, W. W. 1993c. What in the world do we hear? An ecological approach to auditory source perception. Ecological Psychology 5(1), 1-29.

Gaver, W. W., Smith, R. B. & O'Shea T. 1991. Effective sounds in complex systems: the ARKOLA simulation. Teoksessa Proceedings of the Conference on Human Factors in Computer Systems, New Orleans, Louisiana, USA, April 28-May 2, 1991, Reading, MA, USA: ACM Press/Addison-Wesley, 85-90.

Gibson, J. J. 1966. The senses considered as perceptual systems. Boston: Houghton Mifflin.

Goldstein, E. B. 1999. Sensation & perception. 5. painos. USA: Brooks/Cole Publishing Company.

Goose, S. & Möller, C. 1999. A 3D audio only interactive Web browser: using spatialization to convey hypermedia document structure. Teoksessa Proceedings of the seventh ACM International Conference on Multimedia, Orlando, Florida, USA, Oct. 30-Nov. 5, 1999, New York: ACM Press, 363-371.

Hereford, J & Winn, W. 1994. Non-speech sound in human-computer interaction: A review and design guidelines. Journal of Educational Computing Research, 11(3), 211-233.

Howard, J. H., Jr., & Ballas, J. A. 1980. Syntactic and semantic factors in the classification of nonspeech transient patterns. Perception & Psychophysics, 28(5), 431-439.

Huopaniemi, J. 1999. Virtual acoustics and 3-D sound in multimedia signal processing. Helsingin Teknillinen korkeakoulu. Sähkö- ja tietoliikennetekniikan osasto, akustiikan ja äänenkäsittelytekniikan laboratorio. Väitöskirja.

Jauhiainen, T. 1995. Kuulo ja viestintä. Helsinki: Yliopistopaino.

Jones, D. 1993. Objects, streams, and threads of auditory attention. Teoksessa A. Baddeley & L. Weiskrantz (toim.) Attention: Selection, awareness, and control. A tribute to Donald Broadbent, Oxford: Clarendon Press, 87-104.

Jot, J.-M. 1997. Real-time spatial processing of sounds for music, multimedia and interactive human-computer interfaces. Multimedia Systems, 7(1), 55-69.

Kendall, G. 1995. A 3-D sound primer: directional hearing and stereo reproduction. [viitattu 13.5.2002]. Saatavilla www-muodossa <http://www.nwu.edu/musicschool/classes/3D/pages/sndPrmGK.html>.

Kleiner, M., Dalenbeck, B.-I. & Svensson, P. 1993. Auralization - an overview. Journal of the Audio Engineering Society, 41(11), 861-875.

Kobayashi, M., Schmandt, C. 1997. Dynamic Soundscape: mapping time to space for audio browsing. Teoksessa Conference Proceedings on Human factors in computing systems, Atlanta, GA, USA, March 22-27, 1997, New York: ACM Press, 194-201.

Kramer, G. 1994. An introduction to auditory display. Teoksessa G. Kramer (toim.) Auditory Display: Sonification, Audification and Auditory interfaces, Reading, MA, USA: Addison-Wesley Publishing Company, 1-77.

Laurel, B. 1991. Computers as theatre. Reading, MA, USA: Addison-Wesley Publishing Company.

Mansur, D. L., Bly S. A., Frysinger, S. P , Lunney, D, Metzrich, J. J. & Morrison, R. C. 1985. Communication with sound (panel session). Teoksessa W. Buxton (toim.) Proceedings on Human Factors in Computing Systems, San Francisco, CA, USA, 1985, New York: ACM Press, 115-119.

McGrath, D. 1995. High resolution simulation of acoustic environments. [viitattu 13.5.2002]. Saatavilla www-muodossa <http://www.headwize.com/tech/lake1_tech.htm>.

Mereu, S. W. & Kazman, R. 1996. Audio enhanced 3D interfaces for visually impaired users. Teoksessa Proceedings on Human Factors in Computing Systems, Vancouver, British Columbia, Canada, Apr. 13-18, 1996, New York: ACM Press, 72-78.

Mynatt, E. D. & Weber, G. 1994. Nonvisual presentation of graphical user interfaces: contrasting two approaches. Teoksessa Proceedings of the Conference on Human factors in computing systems, Boston, USA, Apr. 24-28, 1994, New York: ACM Press, 166-172.

Perrott, D. R., Saberi, K., Brown, K. & Strybel, T. Z. 1990. Auditory psychomotor coordination and visual search performance. Perception & Psychophysics, 48(3), 214-226.

Saue, S. 2000. A model for interaction in exploratory sonification displays. Teoksessa Proceedings of the International Conference on Auditory Display, Georgia Institute of Technology Atlanta, Georgia, USA, Apr. 2-5, 2000, 1-5.

Sawhney, N. & Schmandt, C. 1997. Design of spatialized audio in nomadic environments. Teoksessa Proceedings of the International Conference on Auditory Display, Palo Alto, CA, USA, Nov. 2-5, 1997, 109-113.

Scaletti, C. & Craig, A. B. 1991. Using sound to extract meaning from complex data. Teoksessa Proceedings SPIE, Vol. 1459, 207-219.

Schaeffer, P. 1967. Traité des objets musicaux. Uudistettu painos. Paris: Seuil.

Schafer, R. M. 1977. The tuning of the world. New York: Knopf.

Schmandt, C. & Mullins, A. 1995. AudioStreamer: exploiting simultaneity for listening. Teoksessa Proceedings of the Conference on Human Factors and Computing Systems, Denver, Colorado, USA, 1995, New York: ACM Press, 218-219.

Smith, S., Pickett, R. M. & Williams, M. G. 1994. Environments for exploring auditory representations of multidimensional data. Teoksessa G. Kramer (toim.) Auditory Display: Sonification, Audification and Auditory interfaces, Reading, MA, USA: Addison-Wesley Publishing Company, 167-183.

Tarasti, E. 1990. Johdatusta semiotiikkaan: esseitä taiteen ja kulttuurin merkkijärjestelmistä. Helsinki: Gaudeamus.

Vanderveer, N. J. 1979. Ecological acoustics: human perception of environmental sounds. Dissertation Abstracts International. 40/09B, 4543. University Microfilms No. 8004002.

Walker, A. & Brewster, S. A. 2000. Spatial audio in small screen device displays. Personal Technologies, 4(2), 1-14.

Walker, R. 1987. The effects of culture, environment, age, and musical training on choices of visual metaphors for sound. Perception & Psychophysics, 42(5), 491-502.

Warren, W. & Verbrugge, R. 1984. Auditory perception of breaking and bouncing events: a case study in ecological acoustics. Journal of Experimental Psychology: Human Perception and Performance, 10(5), 704-712.

Wenzel, E. M. 1992. Localization in virtual acoustic displays. Presence: Teleoperators and Virtual Environments, 1(1), 80-107.

Wenzel, E. M. 1994. Spatial sound and sonification. Teoksessa G. Kramer (toim.) Auditory Display: Sonification, Audification and Auditory interfaces, Reading, MA, USA: Addison-Wesley Publishing Company, 127-150.

Wenzel, E. M., Wightman, F. & Kistler, D. J. 1991. Localization with non-individualized virtual acoustic display cues. Teoksessa Proceedings of the Conference on Human Factors in Computing Systems, New Orleans, Louisiana, USA, Apr. 27-May 2, 1991, New York: ACM Press, 351-359.

Williams, S. M. 1994. Perceptual principles in sound grouping. Teoksessa G. Kramer (toim.) Auditory Display: Sonification, Audification and Auditory interfaces, Reading, MA, USA: Addison-Wesley Publishing Company, 95-125.

Wu, J.-R., Duh C.-D., Ouhyoung, M. & Wu, J.-T. 1997. Head motion and latency compensation on localization of 3D sound in virtual reality. Teoksessa Proceedings of the ACM Symposium on Virtual Reality Software and Technology, Lausanne, Switzerland, Sept. 15-17, 1997, New York: ACM Press, 15-20.