

# **Multimodaalisuus käyttöliittymäsuunnittelijan näkökulmasta**

Graafisen ja puhekäyttöliittymän symbioosi

Matti Kanninen  
Taideteollinen korkeakoulu  
Medialaboratorio  
MA in New Media  
Lopputyö 23.4.2003

## **Alkusanat**

Nyt kun työ on vihdoin valmis, haluan kiittää sen tekemisessä avustaneita henkilöitä.

Suuret kiitokset Medialaboratorion lopputyöseminaarin väelle. Erityisesti haluaisin kiittää seminaarin vetäjiä Heidi Tikkaa ja Maria Koskijokea henkilökohtaisesta keskusteluavusta ja kannustuksesta sekä opponentejani Johanna Höysniemeä ja Päivi Romppasta erinomaisista kommentteista ja keskusteluista. Erityisesti Päivi ylitti opponentin velvollisuutensa toimittamalla toistuvasti tietoja, joita ilman tämä työ ei ehkä olisi vielä valmistunut.

Haluan myös kiittää työni ohjaajaa Kristiina Jokista, joka johdatti minut tiedon lähteille ja antoi keskusteluapua työn edetessä.

Lisäksi haluan kiittää vanhempiani Jukka ja Kaija Kannista keskusteluista ja työn kommentoinnista. Äitiäni haluan lisäksi kiittää erittäin paljon työn rakenteen kommentoinnista ja kirjoitusasun uutterasta tarkastamisesta.

Matti Kanninen

# Sisällysluettelo

<b>SISÄLLYSLUETTELO .....</b>	<b>1</b>
<b>1. TAUSTA.....</b>	<b>4</b>
1.1 TYÖN TAUSTA.....	4
1.2 OMA TAUSTA.....	4
<b>2. TYÖN TARKOITUS JA RAKENNE .....</b>	<b>6</b>
2.1 TYÖN TARKOITUS.....	6
2.2 TYÖN RAKENNE .....	7
<b>3. JOHDANTO.....</b>	<b>9</b>
3.1 MITÄ MULTIMODAAALISUUS ON? .....	9
3.2 MULTIMODAAALISUUDEN JA MULTIMEDIAN SUHDE .....	13
3.3 MIKSI MULTIMODAAALISET KÄYTTÖLIITTYMÄT? .....	14
3.4 GRAAFINEN KÄYTTÖLIITTYMÄ .....	18
3.5 PUHEKÄYTTÖLIITTYMÄ .....	19
3.6 TYÖN SUHDE TEKNIikkaAN .....	21
3.6.1 Käyttäjakeskeinen vs. järjestelmäkeskeinen näkökulma ....	22
3.7 ESIMERKKISOVELLUKSET.....	23
3.7.1 CarMMI .....	24
3.7.2 LARRI .....	25
3.7.3 Interact.....	27
<b>4. SUUNNITTELIJAN MAAILMA .....</b>	<b>30</b>
4.1 IHMISET OVAT ERILAISIA.....	30
4.1.1 Ihminen koneena .....	31
4.1.2 Fyysiset erot.....	32
4.1.3 Psyykkiset erot .....	32
4.1.4 Kulttuuriset erot.....	33
4.1.5 Kokemuksen merkitys.....	33
4.1.6 Taustatieto ja mentaalimallit.....	34
4.1.7 Muuttuvat ominaisuudet.....	35
4.1.8 Motivaation merkitys .....	35
4.2 KÄYTTÖTILANTEET OVAT ERILAISIA .....	36
4.3 MONIULOTTEINEN MAALI .....	37

4.4	RATKAISUTAPOJA.....	39
4.5	SUUNNITTELIJAN TYÖKALUT.....	40
4.6	RAKENNUSPALIKAT.....	43
4.7	SYÖTE- JA TULOSTUSLAITTEET .....	44
4.8	RAHAMAAILMAN RAJOITUKSET .....	46
4.8.1	Monimutkaisuuden seuraukset .....	46
4.8.2	Markkinoinnin haasteet .....	48
4.8.3	Multimodaalisuuden mahdollisuudet.....	49
<b>5.</b>	<b>MODALITEETTIEN OMINAISUUDET .....</b>	<b>51</b>
5.1	AFFORDANSSIT, RAJOITUKSET JA KONVENTIOT .....	51
5.2	TIEDON AJALLINEN KESTO .....	52
5.3	RINNAKKAINEN VS. SARJAMUOTOINEN TIETO .....	53
5.4	TUNNISTAMINEN VS. MUISTAMINEN .....	55
5.5	SUUNNITELMALLINEN VS. TILANNEKOHTAINEN TOIMINTATAPA .....	57
5.6	GRAAFINEN JA PUHEKÄYTTÖLIITTYMÄ TUKEVAT TOISIAAN.....	58
5.7	PUHEMODALITEETTIN OMINAISUUDET .....	59
5.8	PUHEKÄYTTÖLIITTYMIEN TEKNISET ERITYISPIIRTEET .....	62
5.8.1	Puheentunnistuksen rajoitukset.....	62
5.8.2	Puheentuoton rajoitukset .....	64
5.8.3	Dialogin hallinnan malleja.....	65
<b>6.</b>	<b>MODALITEETTIEN VALINTA JA YHDISTÄMINEN.....</b>	<b>67</b>
6.1	VUOROVAIKUTUSTILANTEEN MALLI.....	69
6.2	YHDISTÄMISEN TAVAT .....	71
6.2.1	Toisiaan täydentäen (Complementarity) .....	71
6.2.2	Toisiaan vahvistaen (Redundancy).....	72
6.2.3	Tasa-arvoisesti (Equivalence).....	72
6.2.4	Erikoistuneesti (Specialisation).....	72
6.2.5	Yhtäaikaisesti (Concurrency).....	73
6.2.6	Modaliteettia muuttaen (Transfer) .....	73
6.2.7	Peräkkäin vs. rinnakkain .....	73
6.3	MODALITEETTIEN TAHDISTUS.....	74
6.4	KÄYTTÄJÄN MUKAAN .....	76
6.4.1	Tuottomodaliteetit .....	76
6.4.2	Vastaanottomodaliteetit .....	78
6.4.3	Eriyisryhmät .....	79

6.4.4	Design for all .....	80
6.5	KÄYTTÖTILANTEEN MUKAAN.....	80
6.5.1	Tuottomodaliteetit .....	81
6.5.2	Vastaanottomodaliteetit .....	82
6.5.3	Paikka- ja kontekstiriippuvuus .....	83
6.5.4	Multimodaalisuus ja turvallisuus .....	85
6.6	MODALITEETIN VALINTA SISÄLLÖN JA TEHTÄVÄN PERUSTEELLA .....	85
6.6.1	Tuottomodaliteetit .....	87
6.6.2	Vastaanottomodaliteetit .....	89
6.6.3	Tehtävän sisäinen tilanneriippuvuus.....	91
6.7	MUUTA YHDISTÄMISESSÄ HUOMIOITAVAA.....	92
6.7.1	Konsistenssi modaliteettien välillä .....	92
6.7.2	Painotus (visuaalis- vai puhepainotteinen).....	92
<b>7.</b>	<b>IHMISEN JA KONEEN KOHTAAMINEN .....</b>	<b>94</b>
7.1	EPÄSYMMETRINEN KÄYTTÖTILANNE .....	94
7.2	TYÖKALU VS. TOIMIJA .....	95
7.3	PUHEEN ERITYISASEMA .....	97
<b>8.</b>	<b>LOPPUPÄÄTELMÄT .....</b>	<b>100</b>
	<b>LÄHDELUETTELO .....</b>	<b>103</b>

# 1. Tausta

## 1.1 Työn tausta

Sain ensimmäisen siemenen aiheelle muutama vuosi sitten, kun satumalta juttelin vanhan miehen kanssa raitiovaunussa. Jossain vaiheessa keskustelua mies alkoi hieman tuhtuneena kertoa, kuinka hankalaa nykyaikaisia pankkiautomaatteja on käyttää. Näyttö on pieni ja epäselvä vanhoille silmille eikä kommunikointi koneen kanssa ole muutenkaan kovin luontevaa. Myöhemmin olin työni kautta mukana eräässä puhekäyttöliittymäprojektissa. Jossain vaiheessa projektia muistin vanhan miehen automaattikäyttöongelmat ja syntyi ajatus graafiseen käyttöliittymään integroidusta puhepohjaisesta avustejärjestelmästä. *Avustejärjestelmä toimisi saumattomasti graafisen sovelluksen kanssa ja mahdollistaisi vaikeiden kohtien täyttämisen vuorovaikutteisesti puheen avulla.*

Ajatus jäi muhimaan ja noin vuosi sitten ajattelin kehittää ideaa eteenpäin lopputyön merkeissä. Viime keväänä esittelin aihetta lopputyöseminaarissa ja pohdittuani tarkemmin tajusin, että aihe on monisyisempi kuin ensimmäisellä oivalluksen hetkellä pystyin ymmärtämään. Erinäisten keskustelujen jälkeen alkoi muovautumaan ajatus aiheen yleisemmästä tarkastelusta. Tämän lopputyön kautta yritän saada itselleni selvemmän kuvan aiheesta ja samalla toivottavasti hyödyttää muita aiheesta kiinnostuneita.

## 1.2 Oma tausta

Olen pienestä asti ollut kiinnostunut tekniikasta. Olen viettänyt lapsuuteni sinä aikana kun kotitietokoneet ovat tulleet Suomessa laajemmin markkinoille Commodore VIC-20:n ja C-64:n myötä ja olen ollut itsekin alan harrastaja.

Koulutustaustani on myös tekninen. Olen opiskellut useita vuosia Helsingin teknillisessä korkeakoulussa, viime vuosina tosin vähemmän Medialaboratorion opintojen ja työnteon takia. TKK:lla mielenkiintoisimmat kurssit liittyivät käyttöliittymäsuunnitteluun, työpsykologiaan ja uuden median sisällöntuotantoon.

Työtehtäviini on kuulunut usean vuoden ajan graafisten käyttöliittymien suunnittelu ja toteuttaminen. Aluksi suunnittelu tapahtui enemmän intuition pohjalta yhdessä käyttäjien kanssa, myöhemmin mukaan on tullut suunnitelmallisuus suunnitteluprosessin ja todennettavuus käyttöliittymätestauksen myötä. Myös puhekkäyttöliittymät ovat tuulleet jonkin verran tutuiksi työni puolesta.

Medialaboratorion panos on ollut avata uusia näkökulmia asioiden tekemiseen, erityisesti käyttäjän ja käyttökulttuurin huomioimiseen suunnittelussa.

## 2. Työn tarkoitus ja rakenne

### 2.1 Työn tarkoitus

Lopputyön tarkoituksena on tutustua multimodaalisten käyttöliittymien maailmaan ja miettiä käyttöliittymäsuunnittelijan näkökulmasta, mitä erityiskysymyksiä multimodaalisuus tuo mukanaan sekä käyttäjälle että suunnittelijalle. Tehtävää on rajattu siten, että siinä keskitytään kahteen modaliteettiin, auditiiviseen eli puhe- ja kuuloaistin kautta tapahtuvaan vuorovaikutukseen ja visuaaliseen eli näköaistin kautta tapahtuvaan vuorovaikutukseen, unohtamatta viittauksia muihin modaliteetteihin sopivissa kohdissa. Auditiivinen ja visuaalinen modaliteetti on valittu tarkastelun alle siksi, että alkupe-  
räinen idea pankkisovelluksesta perustui pitkälle niihin ja toisaalta siksi, että ne ovat luonteeltaan melko erilaisia, jolloin niiden yhdistelmistä voi tulla mielenkiintoisia. Mikäli tarkoitus olisi suunnitella käytännön sovellus todelliseen tarpeeseen, tulisi modaliteetit tietysti valita sovelluksen käyttäjien ja käyttötarkoituksen mukaan.

Kummallakin modaliteetilla on omat vahvat puolensa ja ne soveltuvat yksinään käytettyinä erilaisiin käyttöliittymäratkaisuihin. Työn kantavana ajatuksena on, että yhdessä käytettynä modaliteetit mahdollistavat käyttöliittymäratkaisuja, jotka ovat käyttäjän ja käyttötilanteen kannalta toimivampia kuin kummankaan modaliteetin yksinään mahdollistamat ratkaisut.

Keskeisiä kysymyksiä ovat:

- Miten modaliteetit voidaan yhdistää?
- Miten käyttäjä kokee multimodaalisuuden?
- Mitä multimodaalisuus merkitsee käyttöliittymäsuunnittelijalle?



## **2.2 Työn rakenne**

Luvussa 1. kerrotaan tausta ja motivaatio työn tekemiselle sekä tekijän oma tausta työn suhteen.

Luvussa 2. kerrotaan työn tarkoitus ja rakenne.

Luvussa 3. esitellään työn keskeiset käsitteet, pohditaan multimodaalisuuden hyötyjä, selvitetään työn suhde tekniikkaan ja esitellään muutama esimerkkisovellus.

Luvussa 4. pohditaan aluetta, jolla käyttöliittymäsuunnittelija liikkuu ja mitä uutta multimodaalisuus tuo siihen.

Luvussa 5. esitellään graafisen- ja puhe käyttöliittymän erityisominaisuuksia käyttäjän näkökulmasta sekä tuodaan esiin muutamia keskeisiä teknisiä rajoituksia, joita puhe käyttöliittymä tuo mukanaan.

Luvussa 6. mietitään modaliteettien valinnan ja yhdistämisen problematiikkaa. Luvussa pohditaan, kuinka modaliteetit voidaan yhdistää, millaisille käyttäjille ja minkälaisiin käyttötilanteisiin kumpikin tarkasteltavista modaliteeteista sopii. Luvussa käsitellään myös sisällön ja tehtävän vaikutusta modaliteetin valintaan.

Luvussa 7. pohditaan ihmisen ja tietokoneen suhdetta, mietitään miten multimodaalisuus siihen vaikuttaa. Lisäksi luvussa tarkastellaan puheen erityistä asemaa ihmisen kommunikaatiotapana.

Luvussa 8. esitellään työn pohjalta tehdyt johtopäätökset.

Työn keskeisistä kysymyksistä käyttäjän kokemus tulee esille luvuissa erityisesti 3., 5., 6. ja 7. Suunnittelijan näkökulma näkyy ko-

ko työssä ja erityisesti luvussa 4. Modaliteettien yhdistämisen kannalta tärkeimmät luvut ovat 5. ja 6.

## 3. Johdanto

### 3.1 *Mitä multimodaalisuus on?*

Multimodaalisuus viittaa HCI (Human Computer Interaction) keskustelussa useampaan yhtäaikaiseen kanavaan, joilla ihminen ja tietokone voivat kommunikoida keskenään. Aihealue on aika nuori ja monimuotoinen ja vaikka sitä on tutkittu jo jonkin verran, alan terministö ei ole vielä täysin vakiintunut. Alkuosa multi on helpompi. Se tarkoittaa, että useampi modaliteetti on käytössä tai käytettävissä yhtäaikaa.

Modaliteetista on esitetty erilaisia määrittelyjä. Usein käytetty tulkinta on, että modaliteetti tarkoittaa ihmisen käyttämää kommunikaatiotapaa, joka on läheisessä yhteydessä ihmisen eri aisteihin ja aistielimiin. Visuaalinen modaliteetin kommunikaatio tapahtuu näköaistin kautta silmiä käyttäen, auditiivinen kuuloaistin kautta korvia käyttäen jne. (Schomaker ym. 1995)

Modaliteetti erotetaan usein koodauksesta ja mediasta. Maybury ja Wahlster (1998) ovat määritelleet termit seuraavasti:

- Media = Tiedon esitykseen tai tallennukseen käytetty aineellinen kappale, fyysiset informaation kuljetuskeinot (äänet, liikkeet, luonnolliset kielet)
- Koodaus = Viestinnässä käytetty merkkijärjestelmä
- Modaliteetti = Sisäänpäin tulevan informaation käsittelyyn käytetyt aistit (näkö, kuulo, kosketus, haju, maku, tasapaino)

(Jokinen & Raike 2002)

Mediaa voidaan mallintaa usealla eri tasolla. Matalalla fyysisellä tasolla media voi olla esimerkiksi paineaallot, molekyylit tai sähkömagneettinen säteily (Schomaker ym. 1995). Korkeammalla tasolla

media voi olla jotain rakenteellisempaa kuten äänet, liikkeet tai luonnolliset kielet.

Edellä olevaa määrittelyä käyttäen suomenkielinen matkapuhelin-keskustelu voidaan jaotella esimerkiksi seuraavasti: Paineaallot (media) kuljettavat luonnollista kieltä (media). Suomenkieli (koodaus) sekä prosodia eli äänen sävy, painotus ja rytmi (koodaus) vaikuttavat viestien merkityksen muodostumiseen. Viestien tuottamiseen ja vastaanottamiseen käytetty modaliteetti on auditiivinen.

Toinen käytetty tulkinta on, että media on fyysinen esityskerros, jota tulkitaan jonkun tietyn aistin avulla ja modaliteetti on koodaustapa tietyssä mediassa. Näin saman median kautta välittyvät (ja siten samalla aistilla tulkitut) erityyppiset kommunikaatiotavat voidaan erottaa toisistaan. Tämän tulkinnan mukaan esimerkiksi viittomakieli ja kirjoitettu teksti ovat eri modaliteetteja, vaikka ne kumpikin välittyvät valon avulla ja tulkitaan näköaistilla. (Dybkjær ym. 2002. Jokisen ja Raikeen 2002 mukaan).

Saman tyyppisen tulkinnan esittää Bernsen (2002) modaliteettiteoria (Modality theory) yhteydessä. Teorian mukaan media on informaation fyysinen ilmenemismuoto, jonka ihmiset aisteillaan ja järjestelmät sensoreillaan pystyvät havaitsemaan. Media on siten lähellä psykologiassa käytettyä aistimodaliteetin (sensory modality) käsitettä. Bernsen käyttää esityksellisen modaliteetin (representational modality) käsitettä erottelemaan saman median kautta välittyvät eri tyyppiset vuorovaikutustavat kuten puheen ja merkkiäänet. Luokitteluperusteina ovat vuorovaikutuksen kielellisyys, yhdenmukaisuus todellisuuteen (analogisuus), mielivaltaisuus ja dynaamisuus. Lisäksi teoriassa esiintyy modaliteetin sisäisen informaatiokanavan (information channel) käsite, joka on osittain yhteneväinen aiemmin esiteltyjen tulkintojen koodauksen kanssa.

Kaikilla tulkinnoilla on omat hyvät ja huonot puolensa. Koodauksen sisällyttäminen mukaan modaliteettiin helpottaa tiedon esitystavan ja modaliteettien suhteen arvioimista. Erilaiset koodaustavat ovat kuitenkin usein kulttuurisidonnaisia, keinotekoisia ja niitä on lukematon määrä, jolloin myös niihin perustuvia modaliteetteja on lukematon määrä, mikä hankaloittaa modaliteettien analysointia. Modaliteettien tulkinta yhteneväiseksi aistien kanssa toimii puolestaan arvioitaessa multimodaalisuuden vaikutuksia käyttäjän ja käyttötilanteen fyysisiin ominaisuuksiin. Aisteihin sidotut modaliteetit ovat kuitenkin sen verran korkean tason käsitteitä, etteivät ne sovi sellaisenaan hienojakoiseen analyysiin. Bernsenin tulkinta on jotain kahden edellisen väliltä. Siinä käytetyt luokitteluperusteet ovat yleismaailmallisia ja niiden perusteella modaliteetteja on rajattu määrä.

Tässä työssä käytetään usein tulkintaa, jossa modaliteetti sisältää myös koodauksen eli puhetta pidetään itsenäisenä modaliteettina sen sijaan, että siihen viitattaisiin koodauksena auditiivisessa modaliteetissa. Ne ominaisuudet, jotka ovat aistille ominaisia, vaikuttavat kaikkeen kyseisen aistin kautta välittyvään informaatioon koodauksesta riippumatta. Olennaista on pitää mielessä, että koodaus vaikuttaa havaitsemisprosessiin.

Modaliteetti voi olla viestin kulkusuunnan perusteella vastaanotto- (perception/input) tai tuottotyyppinen (production/output) tai molempia yhtäaikaan (Bernsen 2002). Ellei toisin mainita, kulkusuuntaa ilmaisevia termejä käytetään tässä työssä ihmisen näkökulmasta katsottuna eli vastaanottomodaliteetit tarkoittavat niitä, joiden kautta ihminen voi vastaanottaa informaatiota ja tuottomodaliteetit vastaavasti niitä, joiden kautta ihminen voi tuottaa informaatiota.

Oheisissa taulukoissa (3.1 ja 3.2) on esitetty ihmisen aisteihin liittyvät vastaanotto- ja tuottomodaliteetit sekä vastaavat tietokoneissa käytössä olevat modaliteetit.

*Taulukko 3.1. Vastaanottomodaliteetit*

<b>Aisti</b>	<b>Aistielin (ihminen)</b>	<b>Vastaanotin (kone)</b>	<b>Modaliteetti</b>
Näköaisti	Silmät	Kamera	Visuaalinen
Kuuloaisti	Korvat	Mikrofoni	Auditiivinen
Lihasaisti	Hermot, lihakset	Paikkasensorit, kamera	Kinesteettinen
Tuntoaisti	Iho	Paineherkät alustat	Taktiilinen
Hajuaisti	Nenä	Keinonenä	Olfaktorinen
Makuaisti	Kieli, Nenä	Keinokieli, keinonenä	Gustatoorinen
Tasapainoaisti	Tasapainoelin	Tasapainoanturit	Vestibulaarinen
		MEG, EMG, ENG	(Aivotoiminta)

*Taulukko 3.2. Tuottomodaliteetit*

<b>Modaliteetti</b>	<b>Ihminen</b>		<b>Kone</b>	
	<b>Elimet</b>	<b>Media/Koodaus</b>	<b>Laite</b>	<b>Media/Koodaus</b>
Visuaalinen	Raajat, keho	Eleet, liikkeet	Graafiset näytöt, merkkivalot	Pysähtynyt ja liikkuva kuva, värit
Auditiivinen	Suu, kädet, jalat	Puhe, laulu, äänet, taputus, tömistely	Kaiuttimet, muut mekaaniset värähtelijät	Puhe, laulu, äänet
Kinesteettinen	Raajat, keho	Kosketukset, eleet, liikkeet	Vastavoima (Force feedback) laitteet	Tärinä, vastavoima
Taktiilinen	Raajat, keho	Kosketukset	Värinähiiri	Värinä
Olfaktorinen	Keho	Ominaisuus	Tuoksugeneraattorit	Tuoksut
Vestibulaarinen			Liikkuvat alustat	Kallistukset
(Aivotoiminta)	Aivot	EEG, MEG	Elektrodit	Sähkösignaalit

Taulukoista voidaan huomata, että aistikeskeisestä tulkinnasta johtuen myös tuottomodaliteetit määrittyvät vastaanottomodaliteettien kautta.

Usein käytetään myös termiä haptinen modaliteetti, joka pitää sisällään kaiken liikkeeseen, vartalon ja raajojen asentoihin ja kosketukseen liittyvän kommunikaation. Edellä olevissa taulukoissa haptinen on jaettu taktiiliseen ja kinesteettiseen modaliteettiin.

Kaikkia edellä esitettyjä modaliteetteja voidaan käyttää tänä päivänä ihmisen ja koneen väliseen vuorovaikutukseen. Käytännössä monet modaliteeteista kuten maku ja hajuaistiin liittyvät sekä suoraa aivotoimintaa käyttävät ovat kuitenkin vielä tutkimusasteella tai käytössä vain erikoissovelluksissa.

Aivotoimintaa ei yleensä pidetä itsenäisenä modaliteettina, vaan ajatellaan, että kaikki aivojen kanssa tapahtuva kommunikaatio tapahtuu aistien kautta. Nykyaikaisilla aivomittausmenetelmillä (EEG, MEG) on kuitenkin mahdollista tarkkailla aivotoimintaa suoraan jopa niin tarkasti, että alkeellisia tietokonesovelluksia pystytään ohjaamaan ajatuksella (Lehtonen & Sams 2003).

### ***3.2 Multimodaalisuuden ja multimedian suhde***

Nykyaikaisen multimediatietokoneen käyttö on periaatteessa multimodaalista, koska käyttäjä syöttää tietoa koneeseen näppäimistön (haptinen modaliteetti) ja hiiren (haptinen modaliteetti) avulla ja saa palautetta tietokoneelta näytön (visuaalinen) ja äänijärjestelmän (auditiivinen) kautta.

Multimodaalisten ja multimedija järjestelmien välinen ero ei ole välttämättä ilmiselvä. Molemmissa käytetään useita kanavia tiedon siirtoon käyttäjältä tietokoneeseen ja tietokoneesta käyttäjälle. Tutkijat tekevät eron järjestelmän suorittaman tulkinnan ja tarkastelunäkökulman mukaan. Järjestelmän näkökulmasta multimodaaliset sovellukset pystyvät kommunikoimaan käyttäjän kanssa eri tyyppisiä kanavia käyttäen sekä rakentamaan kommunikaatiosta eri tasoisia abstraktioita. Abstraktioiden avulla järjestelmät voivat automaattisesti tulkita merkityksiä käyttäjän syötteistä ja tuottaa niitä palautteisiin. Multimediajärjestelmiltä automaattisten abstraktioiden rakentamisen kyky puuttuu (Nigay & Coutaz 1993). Esimerkiksi multimediatyöasemassa olevalla videokameralla voidaan tallettaa videokuva koneeseen, editoida sitä ja katsella lopputulosta ruudulta.

Sama järjestelmä muuttuu multimodaaliseksi, jos videokameran avulla voidaan tunnistaa pään liikkeitä ja niiden avulla ohjailta tietokoneen toimintaa. Järjestelmä tulkitsee päänliikkeiden merkityksen tietokoneen ohjauksen komentoiksi. Käyttäjän näkökulmasta myös multimedialjärjestelmän käyttö edellyttää multimodaalista havaitsemista.

### **3.3 Miksi multimodaaliset käyttöliittymät?**

Ihmisen ja tietokoneen välisessä kommunikaatiossa on pitkään rajoitettu muutamaan peruskaavaan. Yleisin käyttötilanne lienee sellainen, jossa käyttäjä istuu tietokoneen ääressä, käyttää näppäimistöä ja hiirtä tietojen syöttämiseen ja saa tietokoneen muodostaman palautteen graafisen näytön välityksellä, mahdollisesti yksinkertaisilla äänipalautteilla höystettynä. Nykypäivänä myös monet arkipäiväiset laitteet, kuten kodinkoneet, viihde-elektronikka ja autot sisältävät tietokoneen muodossa tai toisessa. Puhutaan usein sulautetuista järjestelmistä, joissa tietokone on ikään kuin sulautettu muuhun järjestelmään, eikä käyttäjä ehkä koe käyttävänsä tietokonesovellusta, vaan esimerkiksi matkapuhelinta tai MP3-soitinta itsenäisinä tuotteina. Näidenkin laitteiden käyttöliittymät perustuvat useimmiten erilaisiin paineltaviin näppäimiin sekä näyttöihin ja merkkivaloihin.

Viestintäasiantuntijat tuovat usein esille, että ihmisten välisessä kommunikaatiossa suurin osa viestin merkityksestä välittyy muiden kuin sanallisten viestien kautta. Äänensävy, painotukset, tauot, kasvonilmeet, käsien liikkeet, vartalon asennot ja muut sanattomat viestit välittävät paljon informaatiota. Osa sanattomista viesteistä saattaa tukea sanallisia viestejä, osa saattaa olla niiden kanssa ristiriidassa ja osa voi olla jopa sellaisia, että niitä on vaikea välittää sanallisesti.

Multimodaaliset käyttöliittymät ovat yritys laajentaa ihmisen ja koneen välistä vuorovaikutusta. Ajatellaan, että lisäämällä mahdollisia



vuorovaikutuskanavia, voidaan vuorovaikutusta jollain tavalla parantaa. Suuri osa alan tutkimuksesta perustuu ihmisten välisen kommunikaation tutkimukseen, mallintamiseen ja mallien tekniseen toteutukseen, mistä johtuen myös monet olemassa olevat sovellukset nojaavat ajatukseen, että mitä ihmismäisempi sovelluksen vuorovaikutustapa on, sitä luonnollisempi se on ja sitä kautta parempi käyttäjälle. Ihmismäinen kone on kuitenkin ongelmallinen monessa mielessä.

On turha kuitenkaan rajoittua pelkästään ihmisten vuorovaikutuksen imitointiin. Eiväthän nykyisetkään näyttöön, näppäimistöön ja hiireen perustuvat sovellukset ole kovin luonnollisia, mutta silti ihmiset ovat oppineet niitä käyttämään ja saavat niillä paljon aikaan. Samalla tavalla voidaan ajatella, että tietokonetta on suhteellisen helppo oppia käyttämään myös muita vuorovaikutuskanavia käyttäen, jos käyttöliittymä on suunniteltu tarpeeksi helpoksi oppia ja käyttää. (Jokinen & Raike 2002) (Toisaalta Internetin suosikkisovellusten, kuten sähköpostin, pikaviestinnän ja chattien, viehätys perustuu luultavasti osittain siihen, että osa inhimillisen kommunikaation osatekijöistä suodattuu pois.)

Nykyinen teknologia on vielä lapsenkengissään, jos ajatellaan, että tavoitteena on kotihumanoidi, jonka kanssa voit sujuvasti keskustella ja joka ymmärtää olla hiljaa äänensävyä ja kasvonilmeitä perusteella silloin, kun tulet kotiin huonotuulisena ja väsyneenä rankan työpäivän jälkeen. Multimodaalisilla käyttöliittymillä on kuitenkin paljon annettavaa jo ennen kuin edellä kuvattu visio on todellisuutta.

Tähän mennessä tietokoneen käyttö on tapahtunut useimmiten toimistopöydän tai sitä vastaavan työpisteen ääressä. On kuitenkin paljon käyttötilanteita, jotka poikkeavat huomattavasti perinteisestä toimistoympäristöstä. Yhä useammat ihmiset käyttävät pieniä tas-

kutietokoneita (PDA, Personal Digital Assistant) ja matkapuhelimiin on tullut mahdollisuus ladata kehittyneitä sovelluksia, mistä johtuen tietokonesovelluksen käyttö tapahtuu enenevässä määrin käyttäjän ollessa liikkeellä. Liikkuessa ei useinkaan ole mahdollista käyttää esimerkiksi käsiä samalla tavalla kuin toimistopöydän ääressä, jolloin kommunikaatio täytyy hoitaa jotain muuta modaliteettia, esimerkiksi puhetta, käyttäen. Vastaavanlainen tilanne on myös autossa, kun kädet ja katse täytyy varata ajamiseen tai erilaisissa korjaustehtävissä, joissa kädet on sidottu tehtävän suorittamiseen.

Vaihtelevan käyttöympäristön lisäksi pienille tietokonelaitteille on ominaista myös rajoittuneet tiedonsyöttömahdollisuudet (pieni näppäimistö, ei ollenkaan näppäimistöä, ei hiirtä) ja pienikokoinen näyttö. Voidaan ajatella, että tällaisten laitteiden käyttöä voitaisiin helpottaa käyttötilanteesta riippumatta, lisäämällä erilaisia vuorovaikutustapoja. Muita potentiaalisia multimodaalisuuden käyttökohteita ovat enenevässä määrin kaikkialle laajenevan tietokoneistamisen (pervasive, ubiquitous computing) mukanaan tuomat sovellukset, kuten esimerkiksi puettavat tietokoneet, älykkäät tilat ja robottien kanssa kommunikointi.

Paitsi käyttötilanteet, myös käyttäjät ovat erilaisia. On monia ihmisryhmiä, jotka eivät pysty täysipainoisesti hyödyntämään tämän päivän teknologiaa, koska se on suunniteltu valtaväestön tarpeista lähtien. Näkövammaisille ja heikkonäköisille näytöt ovat ongelmallisia, erityisesti pienet näytöt kannettavissa laitteissa. Sairauksien tai vammautumisten takia käsien käyttö on hankalaa tai mahdotonta joillekin ihmisille. Visuaalisen tiedon omaksuminen voi olla vaikeaa lukihäiriöisille.

Väestön ikärakenne vanhenee koko ajan ja vanhetessa ihmisen aistit ja motoriset kyvyt heikkenevät samoin kuin sairauksien määrä lisääntyy, joten yhä useampi ihminen kuuluu tulevaisuudessa eri-

tyisryhmien piiriin. Multimodaalisten käyttöliittymien avulla voidaan helpottaa erityisryhmien tietokoneen käyttöä. Toisaalta "Design for all" tai "Universal design" ajattelun mukaisesti erityisryhmien lähtökohdista suunnitellut sovellukset saattavat olla helppokäyttöisempiä myös valtaväestölle.

Blattner ja Dannberg (1996) luettelevat seuraavia hyötyjä, joita multimodaalisilla käyttöliittymillä voidaan saavuttaa:

- Modaliteettien synergia: Modaliteetit voivat tukea toisiaan sekä syötteen tulkinnassa että palautteen antamisessa. Useamman modaliteetin käyttö voi esimerkiksi parantaa tunnistuksen tarkkuutta.
- Eri modaliteettien hyödyt. Eri modaliteeteilla on niille tunnusomaiset ominaisuudet. Yhdistämällä eri modaliteettien parhaat ominaisuudet voidaan saada aikaan parempi käyttöliittymä kuin yksittäisellä modaliteetilla olisi mahdollista. Esimerkiksi objektien valinta tai osoittaminen voi olla hankalaa puhe käyttöliittymän avulla, toisin kuin graafisessa käyttöliittymässä, jossa osoittaminen ja valinta ovat perustoimintoja.
- Uudet sovellukset: Uusia sovelluksia uusille käyttäjille, uusiin käyttöympäristöihin ja -tilanteisiin.
- Valinnan vapaus: Ihmiset voivat valita haluamansa modaliteetin mieltymyksensä ja tilanteen mukaan. Ihmisille, joille joidenkin tiettyjen modaliteettien käyttö on vaikeaa tai mahdotonta vammaisuuden tai sairauden takia, luo modaliteettivalikoiman laajeneminen uusia mahdollisuuksia käyttää sovelluksia.

- Luonnollisuus: Ihmisillä on parempi mahdollisuus käyttää ihmisten välisessä viestinnässä opittuja taitoja ihmisen ja koneen välisessä vuorovaikutuksessa.
- Ympäristöön sopeutuminen: Useamman modaliteetin käyttö mahdollistaa modaliteetin valinnan vaihtuvan tai muuttuvan käyttöympäristön mukaan.

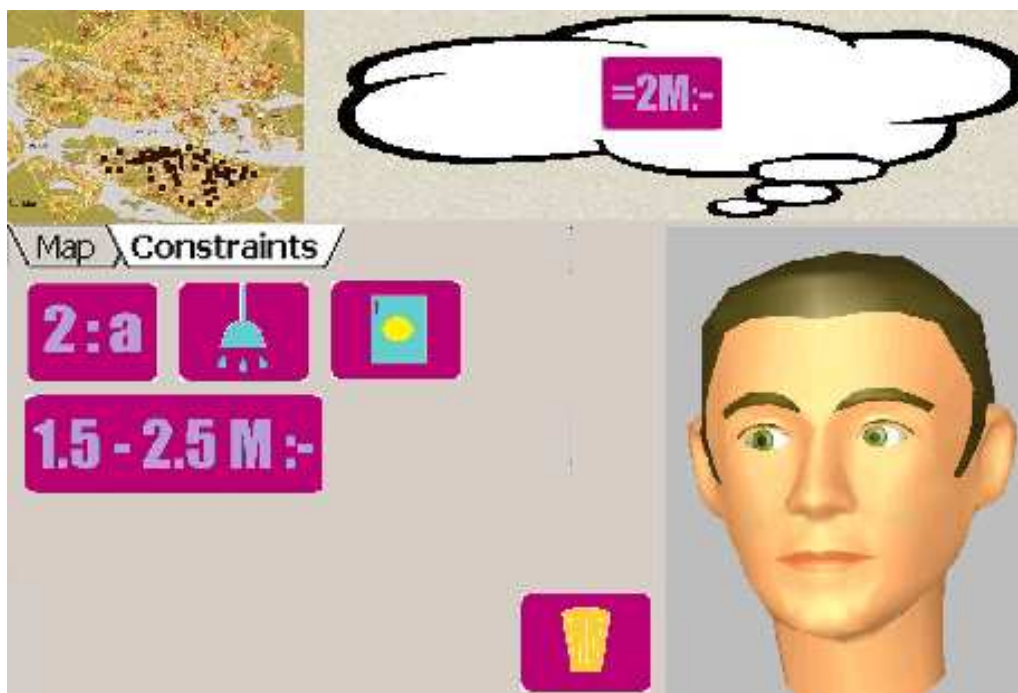
Blattnerin ja Dannbergin mukaan multimodaalisuuden käyttöön liittyy myös ilmiselviä ongelmia. Useamman modaliteetin myötä järjestelmän monimutkaisuus kasvaa. Monimutkaisuuden myötä myös suunnittelutyön määrä kasvaa. Ei myöskään ole olemassa vielä yleistä multimodaalisten järjestelmien viitekehystä, jonka pohjalta voisi arvioida, miten multimodaalisuutta tulisi hyödyntää, mihin sovelluksiin se sopii, mitä modaliteetteja tulisi käyttää ja miten niitä tulisi käyttää. (Gibbon ym. 2000:106)

### **3.4 Graafinen käyttöliittymä**

Graafinen käyttöliittymä mielletään usein Macintosh- tai Windows-tyyppiseksi WIMP-käyttöliittymäksi (WIMP = Windows, Icons, Menus and Pointing Device), jossa hiirellä ohjailten käytetään ikkunoituja sovelluksia. Tässä työssä graafinen käyttöliittymä käsitetään laajemmin ja se voi olla käyttöliittymä, jossa graafisen näytön avulla esitetään mitä tahansa käyttäjän vuorovaikutukseen vaikuttavaa informaatiota.

Esimerkkinä erilaisesta informaatiosta on multimodaalisen vuorovaikutuksen tutkimuksessa käytetyt keinopäät. Keinopää on tietokonegrafiikan avulla ihmisen päätä ja kasvoja jäljittelevä sovellus, jota voidaan ohjata niin, että sen suu, silmät ja muut kasvojen osat liikkuvat samaan tapaan kuin ihmisellä. Keinopään avulla voidaan esimerkiksi tutkia, kuinka paljon puhujan suun liikkeet vaikuttavat kuuntelijan kykyyn ymmärtää puhetta. Toisaalta keinopään avulla

voidaan myös parantaa puheen ymmärtämistä häiriöisessä viestintätilanteessa. Esimerkiksi puhesynteesiä parannetaan antamalla kuuntelijalle visuaalisia vihjeitä syntetisaattorin puhumista äänneistä keinopään avulla.



Kuva 3.1. Keinopää multimodaalisessa sovelluksessa (Gustafson ym. 2002).

Mahdollisuus osoittaa suoraan kohteita näytöllä on merkittävä etu graafisissa käyttöliittymissä. Koska osoittaminen on olennainen osa graafista käyttöliittymää, sitä ei rajata pois tämän työn sisällöstä. Osoittaminen tuo myös yleensä mukanaan uuden modaliteetin, jota ei kuitenkaan erityisesti käsitellä, koska aiheen kunnollinen käsittely vaatisi liikaa resursseja. Perinteisesti osoittaminen tehdään hiirellä, mutta tässä työssä osoitinlaitteeseen ei oteta kantaa, se voi olla yhtä hyvin kosketusnäyttö kuin katseentunnistinkin.

### 3.5 Puhikäyttöliittymä

Puhikäyttöliittymien kehitys on alkanut jo 1700-luvun loppupuolella, kun tehtiin ensimmäiset kokeet tuottaa puheen kaltaisia äänneitä

ihmisen äänentuottoelimiä matkivilla mekaanisilla syntetisaattoreilla. Siitä lähtien kehitys on edennyt tietokoneiden, tietojenkäsittelyn ja kielentutkimuksen kehityksen mukana. Kaupallisiin sovelluksiin puhekäyttöliittymät tulivat mukaan ensin puhesynteesin kautta ja 80-luvun loppupuolelta lähtien myös puheentunnistussovellusten kautta. (Weinschenk & Barker 2000)

Science-fiktio kirjat ja elokuvat lienevät ensimmäisiä lähteitä, joiden kautta suuri yleisö on saanut mielikuvia puhekäyttöliittymistä. Avaruusseikkailu 2001:n keskusteleva keskustietokone HAL on varmaan yksi kuuluisimmista puhekäyttöliittymäkonsepteista. Monissa tarinoissa juttu on viety vielä pidemmälle niin, että kyborgit ovat myös fyysisesti ihmisen näköisiä, eivät ainoastaan käytä peri inhimillistä kommunikaatiotapaa. Tieteiskirjallisuus on siinä mielessä olennaista multimodaalisten käyttöliittymien kannalta, että se luo ennakkomielikuvia ihmisille.

Markkinoilla olevat kaupalliset puhekäyttöliittymät edustavat muutamaa pääsuuntaa. Eräs ryhmä on kehittyneet automaattiset puhejärjestelmät (APJ, IVR = Interactive Voice Response) eli sovellukset, joita käytetään puhelimen näppäimillä ja jotka antavat palautteen puheena, kuten esimerkiksi matkapuhelinvastaaja. Kehittyneemmissä järjestelmissä sovellusten ohjaamiseen käytetään puhelimen näppäinäänien sijaan puhekomentoja. APJ-puolelta ovat peräisin myös monet hierarkkisiin valikoihin perustuvat ratkaisut, jotka toimivat näppäinäänillä kohtalaisesti, mutta puhekäytössä saattavat olla aika kömpelöitä. Tällä sektorilla on myös tehty tutkimusta erilaisten dialogien suunnittelun, sanaston valinnan jne. suhteen.

Toinen ryhmä on äänikomennoilla ohjattavat ohjelmat tai laitteet. Ryhmään kuuluu tietokonepelejä, puhelimia jne. Komennot ovat useimmiten yksinkertaisia ja niiden avulla voidaan korvata joitain näppäinyhdistelmiä.

Kolmas ryhmä on sanelujärjestelmät, joissa järjestelmän tehtävänä on kääntää vapaata puhetta tekstiksi. Kaupallisia on jonkin verran markkinoilla, mutta mitään suurta suosiota ne eivät ole saavuttaneet. Suurin ongelma lienee nopeuden ja tunnistustarkkuuden saaminen tarpeeksi suureksi niin, että järjestelmä tarjoaisi kilpailuedun tai pääsisi edes tasoihin kymmensormijärjestelmällä kirjoittamisen kanssa.

Neljäs ryhmä on kapean alan erikoissovellukset kuten lääkäreiden tai lakimiesten diagnosointijärjestelmät.

Tavallisen puheen lisäksi myös erityyppisiä näköaistin (visuaalinen puhe, huulilta luku, viittoma-avusteinen huulilталuku (cued speech), viittomakieli) ja tuntoaistin kautta välittyviä puhetapoja voidaan periaatteessa käyttää tietokoneen kanssa kommunikointiin, mutta niiden tarkempi käsittely on tämän työn alueen ulkopuolella.

Tässä työssä puhekäyttöliittymä tarkoittaa tapaa, jolla käyttäjä voi puheen avulla olla tietokoneen kanssa vuorovaikutuksessa. Puhe voi toimia kommunikaatiovälineenä kumpaankin suuntaan tahansa, eikä sen tarvitse välttämättä olla kaksisuuntaista.

### **3.6 Työn suhde tekniikkaan**

Käyttöliittymien perimmäinen tarkoitus, ihmisen ja koneen yhteensovittaminen, määrää, että tekniikka on aina olennainen osa käyttöliittymää. Multimodaalisten käyttöliittymien yhteydessä tekniikan merkitys vain korostuu.

Esimerkiksi erilaisten tunnistinten (puhe, ele, katse) käyttämät algoritmit ovat monimutkaisia ja ne vaativat paljon laskentatehoa verrattuna perinteisten käyttöliittymien näppäintenpainallusten ja hiirenliikkeiden tutkimiseen. Eri tunnistinten kautta välittyvä informaatio

tio täytyy lisäksi yhdistää ja tulkita järkeviksi toiminnoiksi, mikä nostaa teknistä vaatimustasoa entisestään. Vastaavasti esimerkiksi syntetisoidun puheen tuotto tai keinopään animointi reaaliajassa tarjoavat teknisiä haasteita.

Edistyksellisyydestään ja monimutkaisuudestaan huolimatta käytetty tekniikka asettaa paljon rajoituksia ja käytännössä käyttöliittymät täytyy suunnitella siten, että tekniikan rajoitukset haittaisivat mahdollisimman vähän käyttäjää. Tavallisia strategioita haittojen vähentämiseksi ovat käyttäjän ohjaaminen niin, etteivät rajoitukset tule normaalikäytössä vastaan ja toisaalta käyttäjän ohjaaminen mahdollisimman hyvin niissä ongelmatilanteissa, joissa järjestelmän rajoitukset ovat tulleet vastaan.

Joskus ajatellaan, että käyttöliittymäsuunnittelijan olisi hyvä pitää itsensä erossa tekniikasta, ettei tehtyjä ratkaisuja tehtäisi tekniikan, vaan käyttäjän lähtökohdista. Ajatellaan, että mikään ei muutu, jos asiat tehdään olemassa olevan tekniikan ehdoilla. Olen samaa mieltä, että aina ei kannata mennä siitä, mistä aita on matalin, vaan koetella tekniikan rajoja ja nähdä vaivaa käyttäjäystävällisemmän lopputuloksen aikaansaamiseksi. Toisaalta olen myös sitä mieltä, että koska käyttöliittymien suhde tekniikkaan on niin vahva, on suunnittelijalle erittäin hyödyllistä tuntea käytettävissä olevan tekniikan mahdollisuudet niin, että ne voidaan suunnittelun yhteydessä valjastaa mahdollisimman hyvin käyttöön.

### **3.6.1 Käyttäjäkeskeinen vs. järjestelmäkeskeinen näkökulma**

Multimodaalisten käyttöliittymien kohdalla vuorovaikutustilannetta voidaan tarkastella käyttäjän ja järjestelmän näkökulmasta. Kummastakin näkökulmasta multimodaalisuus näkyy lisääntyneinä vuorovaikutuskanavina ihmisen ja koneen välillä, mutta kaikki kanavat eivät välttämättä näy käyttäjälle. Vuorovaikutuskanavat voidaan ja-



kaa aktiivisiin ja passiivisiin sillä perusteella, voiko käyttäjä aktiivisesti käyttää niitä järjestelmän ohjaamiseen.

Esimerkiksi puheentunnistuksen tarkkuutta voidaan parantaa käyttämällä puheentunnistimen kanssa yhtäaikaan suunliikkeiden tunnistusta eli huulitalukua. Voidaan ajatella, että huulitaluku on passiivinen modaaliteetti. Järjestelmän kannalta kysymyksessä on multimodaalinen ja varsin monimutkainen tilanne, mutta käyttäjä ei välttämättä koe tilannetta yksimodaalisesta puheentunnistuksesta poikkeavana, ellei hän esimerkiksi joudu tietoisesti suuntaamaan kasvojaan huulia kuvaavaa kameraa kohti. Vaikkei käyttäjä sitä tiedostaakaan, toisen modaaliteetin olemassaolo parantaa kuitenkin järjestelmän käytettävyyttä, koska tunnistustarkkuuden kasvun myötä virhetilanteiden määrä vähenee. Samassa järjestelmässä voi olla sekä aktiivisia että passiivisia modaaliteetteja, jolloin voidaan puhua sekoitejärjestelmistä (blended). Aktiivisten modaaliteettien kautta on yleensä helpompi tulkita käyttäjän aikeita, passiiviset ovat vastavasti vähemmän käyttäjää häiritseviä. (Oviatt & Cohen 2000a).

Tässä työssä pyritään vuorovaikutustilannetta tarkastelemaan enemmän käyttäjän kuin järjestelmäarkkitehtuurin ja teknisen toteutuksen näkökulmasta ja enemmän aktiivisten kuin passiivisten modaaliteettien kautta. Kannattaa kuitenkin muistaa, että käyttöliittymä ja tekniikka kulkevat käsi kädessä ja ihannetapauksessa ne myös suunnitellaan yhdessä.

### **3.7 Esimerkkisovellukset**

Seuraavaksi esitellään muutamia multimodaalisia sovelluksia, joissa graafinen ja puhekäyttöliittymä ovat olennaisena osana. Esimerkkein avulla voi paremmin hahmottaa minkä tyyppisistä sovelluksista on kysymys. Esimerkkeihin viitataan myöhemmin tässä työssä multimodaalisuuden ilmiöitä käsiteltäessä.

### 3.7.1 CarMMI

CarMMI on autotietokoneen ohjaukseen suunniteltu käyttöliittymäprototyyppi. Prototyyppi on integroitu ajosimulaattoriin ja yhdessä ne toimivat tutkimusympäristönä jatkokehitykselle.

Järjestelmässä on noin 30 toimintoa eri tehtäväalueilta, jotka ovat kommunikointi (esim. matkapuhelut), navigointi (esim. reittiohjeet paikasta toiseen) ja viihde. Näihin liittyy erilaisia tehtäviä kuten valinta listasta näytöllä tai ilman, nimien valinta ja muokkaus, numeerisen tiedon syöttö ja graafisten objektien, kuten karttojen käsittely.

Järjestelmää voi ohjata kolmella eri tavalla: puheella, kädellä näyttävillä eleillä ja graafisen näytön yhteydessä olevilla napeilla ja säätimillä. Vastaavasti järjestelmän antama palaute voi tapahtua joko graafisen näytön, esiäänitettyjen puheviestien tai muiden audiosignaalien avulla.

Graafinen näyttö on värillinen ja sen koko on 5,5 tuumaa. Näytöllä esitetään valikon elementtejä ja ne on järjestetty kaaren muotoon, jotta ne yhdistyisivät paremmin ohjaimena käytettävän pyöritettävän ja painettavan säätimen toimintaan. Valikoissa liikkumiseen käytettävän navigointiohjaimen lisäksi näytön yhteydessä on äänenvoimakkuuden säädin ja kaksi näppäintä, joista toisella voidaan käynnistää puheentunnistus ja toinen toimii Back-näppäimenä valikoissa liikuttaessa (vrt. WWW-selaimet).

Puhekäyttöliittymä perustuu enimmäkseen komentosanoihin ja kiinteästi määriteltyihin lauserakenteisiin. Puheentunnistin tunnistaa noin 500 komentosanaa ja 10000 nimeä. Käyttäjä voi käynnistää tunnituksen joko painamalla näppäintä tai sanomalla "Computer". Käyttöliittymä soittaa äänimerkin, kun se on valmis ottamaan puhekomentoja vastaan, samoin kun se lopettaa kuuntelun. (Neuss 2002)

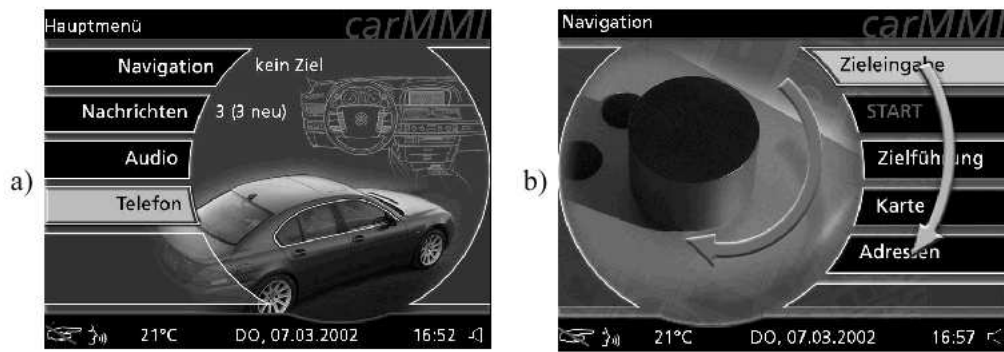


Figure 3. Screenshots of the main menu and a function menu. In the right picture a push-turn-knob has been cut in to point out the correlation between layout and knob rotation.

Kuva 3.2. CarMMI sovelluksen näyttö (Neuss 2002).

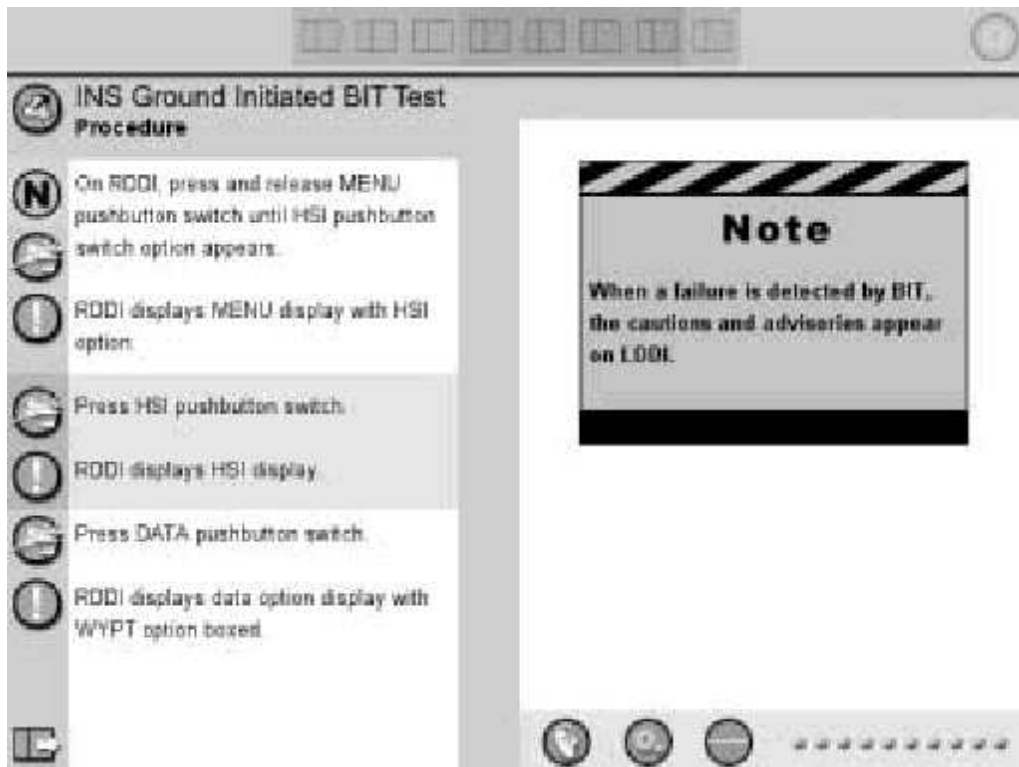
CarMMI kaltaisia käyttöliittymiä on jo olemassa tuotannossa olevissa luksusautoissa. Voidaan ajatella, että ne muidenkin teknisten lisälaitteiden tavoin valuvat halvemman luokan autoihin ajan kuluessa. Tavallisten henkilöautojen lisäksi vastaavia järjestelmiä voisi ajatella käytettäväksi ammattiajoneuvoissa, kuten takseissa, rekoissa sekä erilaisissa työkoneissa.

### 3.7.2 LARRI

LARRI (Language-based Agent for Retrieval of Repair Information) on (hävittäjä)lentokoneiden huoltotöitä avustamaan kehitetty sovellus. Sovellus toimii puettavassa tietokonejärjestelmässä.

Järjestelmää voi ohjata kahdella eri tavalla. Yksinkertaisella valintakiekolla voi valita graafisella näytöllä olevia valikkoelementtejä ja puhetta voi käyttää monipuolisempaan ohjaukseen. Palautetta järjestelmä antaa graafisen näytön ja puheen kautta.

Näyttö on päähän kiinnitettävää mallia ja siinä voidaan esittää valikoiden lisäksi huolto-ohjeita, videoita, animaatioita ja huomautuksia.



Kuva 3.3. LARRI-sovelluksen graafinen näyttö (Bohus 2002).

Puhekäyttöliittymässä on useita käyttötapoja. Käyttäjä voi siirtyä suoria komentoja käyttäen sovelluksen eri osiin. Yksinkertaisten komentojen lisäksi käyttäjä voi käydä pidempiä vuoropuheluja järjestelmän kanssa. Vuoropuheluiden avulla käyttäjä ja järjestelmä voivat yhdessä määrittellä tarpeelliset huoltotehtävät ja tehtävien kuluessa määrittellä tarpeelliset toimenpiteet ja tarkastaa, että ne tulivat tehdyksi. Järjestelmä voi antaa kuhunkin kohtaan tarpeellisen ohjeen puheena, näyttää sen graafisella näytöllä ja lisäksi tarjota lisäinformaatiota kuten animaatioita näytöllä. Toimenpiteen suorituksen jälkeen käyttäjä voi kuitata toiminnon suoritetuksi ja järjestelmä voi tehdä lisäkysymyksiä lopputuloksen varmistamiseksi tai siirtyä seuraavaan tehtävään. Jos käyttäjän työ ei edisty normaali-vauhdissa, voi järjestelmä kysellä käyttäjältä, missä vaiheessa hän on eli järjestelmä voi tehdä tarvittaessa aloitteen keskustelulle. (Bohus 2002)

LARRI on lähellä tämän työn alkuperäisenä siemenenä toiminutta ajatusta puheavusteisesta pankkipalvelusta. Käyttäjä voi suorittaa tehtäviä vuoropuheluna yhdessä järjestelmän kanssa. Pankkisoveluksessa tehtävät olisivat olleet järjestelmän sisäisiä. LARRI:ssa tehtäviin liittyy aina ulkoinen kohde eli käytännössä joku lentokoneen osa tai osajärjestelmä.

### **3.7.3 Interact**

Interact oli usean suomalaisen yliopiston ja yrityksen yhdessä Teke-sin USIX-hankkeessa toteuttama projekti, jonka tuloksena syntyi puhekäyttöinen bussiaikatauluneuvontajärjestelmä. Käyttäjä voi puheen avulla selvittää millä bussilla ja mihin aikaan hän pääsee haluamaansa paikkaan. Sovelluksen käyttö tapahtuu luontevasti puhelimen kautta.

Järjestelmän suunnittelussa on keskitytty erityisesti käyttäjän ja koneen vuorovaikutuksen parantamiseen ja varmaan toimintaan vaihtelevissa käyttötilanteissa. Sovellus mahdollistaa luonnollisen kielen käytön yksinkertaisten komentosanojen sijaan ja pyrkii ohjaamaan keskustelua niin, että se tuntuu käyttäjältä luonnolliselta. Järjestelmässä on myös automaattinen aihealueen tunnistus, jonka avulla käyttäjää voidaan opastaa oikein, jos hän esittää kysymyksiä, jotka eivät kuulu sovelluksen aihealueeseen, kuten matkalippujen hinnat ja muut tyypilliset joukkoliikenneyhtiön neuvontaan esitetyt kysymykset. Eräs projektin tavoitteista oli myös suomenkielisen kieliteknologian edistäminen.

Interactissa on myös pyritty ottamaan erityiskäyttäjät huomioon. Sovellusta voi käyttää puheen lisäksi tekstipohjaisesti näppäimistöä ja siihen on toteutettu viittomakielinen opastus.

Edistyksellisten ominaisuuksien aikaansaamiseksi järjestelmässä on käytetty edistyksellistä tekniikkaa, kuten itseorganisoituvia karttoja

(hermoverkkolaskentamenetelmä) ja agenttipohjaista järjestelmäarkkitehtuuria. (Jokinen ym. 2002)



Kuva 3.4. Syöttö. "Milloin menee seuraava bussi tänne täältä?" (Hurtig 2002).



Kuva 3.5. Palaute. "Bussi 59 menee Pasilaan klo 14:31." (Hurtig 2002).

Bussiaikataulujärjestelmään on tehty multimodaalinen laajennus lisäämällä kosketusnäyttö, jonka avulla käyttäjä voi näyttää kartalta

aikataulukyselyihin liittyviä paikkoja ja viitata niihin puhesyötteissä. Kuvissa 3.4 ja 3.5 on esitetty esimerkit multimodaalisista syöttö- ja palautustilanteista.

Interactin multimodaalinen sovellus perustuu samantyyppiseen modaliteettien yhdistämiseen kuin Richard A. Boltin vuonna 1980 julkaissama koesovellus, jossa käyttäjä pystyi yhdessä kädellä osoittamalla ja puhekomentoja antamalla manipuloimaan mediahuoneen seinälle heijastetulla kartalla olevia geometrisia kuvioita. Boltin artikkelin nimessä oleva "Put-That-There" viittaa tilanteeseen, jossa käyttäjä haluaa siirtää kuviota kartalla. Siirtäminen tapahtuu siten, että käyttäjä osoittaa ensin kuviota kädellään, osoittaa sitten uuden paikan kartalta kädellään ja antaa samalla puhekomennon "Laita tämä tuonne", jossa "tämä" viittaa kuvioon ja "tuonne" uuteen sijoituspaikkaan. (Bolt 1980)

Boltin sovellus toimi lähtölaukauksena multimodaalisten käyttöliittymien kehittämiseksi. Erityisesti siinä käytettyjen osoittamismahdollisuudella varustetun graafisen käyttöliittymän ja puhekäyttöliittymän kyky paikata toistensa puutteita on innostanut tutkijoita (Grasso 1998). Tutkimus on suuntautunut vastaavanlaisiin järjestelmiin, jopa siinä määrin, että esimerkiksi Sharon Oviatt (1999) on kritisoinut muiden multimodaalisten yhdistelmien unohtamista.

## 4. Suunnittelijan maailma

### 4.1 *Ihmiset ovat erilaisia*

Vaikka kaikki ihmiset ovat saman lajin edustajia ja monella tavalla samankaltaisia, on yksilöillä monia ominaisuuksia, jotka erottavat heidät toisistaan. Yksi käyttöliittymäsuunnittelun perusongelma ja perusmotivaatio on se, että ihmiset ovat erilaisia. Muutenhan voitaisiin ajatella, että kun yksi ihminen suunnittelee itselleen sopivan käyttöliittymän, käy se sama kaikille.

Ihmisten erilaisuus tuntuu monien mielestä itsestäänselvyydeltä ja tuntuu hieman naiivilta korostaa sitä tässä yhteydessä. Oman kokemukseni perusteella se ei kuitenkaan ole lainkaan itsestään selvää. Yhä edelleen ohjelmistosuunnittelijat ja koodarit tekevät tuotteita, joiden käyttöliittymän he ovat mallintaneet omasta asiantuntijan näkökulmastaan, joka ei välttämättä avaudu loppukäyttäjille. Yhä edelleen monien yritysten yleisölle tarkoitetut verkkosivut on rakennettu yhtiön sisäisen organisaation pohjalta, joka ei merkitse mitään organisaation ulkopuolisille. Monesti organisaation toimintaa tukemaan tehdään järjestelmiä, jotka on mallinnettu talouselämän lakeihin perustuvien, usein mekaanisten, prosessien mukaan, sen sijaan että ne olisi tehty tukemaan ihmisten jouhevaa yhteistoimintaa. Ajatellaan, että ihmisten tulee oppia ja sopeutua.

Jonkinlaista sopeutumista tapahtuu aina varsinkin, jos motivaatio on tarpeeksi kova, koska ihmiset ovat luonnostaan oppivia ja sopeutuvia. Voidaan kuitenkin ajatella, että sopeutuminen kannattaa tehdä mahdollisimman helpoksi, jottei ihmisten energiaa kulu turhaan tehtävään liittymättömien asioiden hallintaan ja vähentää näin stressiä, jota tuntuu ilmenevän tässä kiireisessä maailmassa muutenkin liikaa.



#### 4.1.1 Ihminen koneena

Yksinkertaistaen ihmisen voi ajatella eräänlaisena orgaanisena koneena. Kehon voi ajatella runkona, toimilaitteina ja voimanlähteenä. Aivot puolestaan voi ajatella muistipankkina ja prosessorina, joka ohjaa ihmisen toimintaa, ajatuksia ja liikkumista.

Ajatus ihmisestä koneena on luultavasti vieras useimmille ihmisille arkikokemuksen perusteella. Ihmiset kokevat itsensä elävinä ja tuntevina ihmisinä, yksilöinä ja osana yhteisöjä, ei koneina. Tällainen pohdinta lähestyy ikäikäistä filosofista pohdintaa siitä, mitä ihmisyyys on ja siihen on turha mennä sen pidemmälle tämän työn puitteissa. Monet tieteet, kuten lääketiede, biologia, psykologia ja kognitiotiede kuitenkin lähestyvät ihmistä konemallin kautta.

Konemalli on siitä hyödyllinen, että sitä kautta voi helpommin ymmärtää ihmisten erilaisuuden. Jokaisen yksilöllinen keho antaa rajoitukset fyysiselle olemiselle ja aivot ohjaavat ihmisen toimintaa ainutkertaisen sisältönsä mukaan. Asiantuntija osaa havaita asioita, joihin maallikko ei kiinnitä mitään huomiota, balettianssija kokee kehonsa eri tavalla kuin "sohvaperuna" ja sokea kuulee ympäristönsä eri tavalla kuin näkevä.

Käytännössä ihmisen kokonaisvaltaista toimintaa on hankala mallintaa ja mitata, joten ihmisten välisiä eroja täytyy jollakin tavalla luokitella niin, että erilaisuutta olisi mahdollista tarkastella käyttöliittymien viitekehysessä. Ben Shneiderman (1998) jakaa eroavaisuudet fyysisten kykyjen lisäksi kognitiivisiin kykyihin, persoonallisuuden eroihin sekä kulttuuriseen ja kansainväliseen vaihtelevuuteen. Hän erottelee vielä erikseen vammaiset ja ikäihmiset. Vammaisuus ja ikä ovat muista luokista eroavia, koska ne voivat liittyä sekä fyysisiin että psyykkisiin kykyihin ja kulttuurisiin eroihin. Jakob Nielsen (1993) luokittelee käyttäjät kokemuksen perusteella, mikä käytän-

nössä tarkoittaa Shneidermanin kulttuurisen luokan erään alaluokan jakamista pienempiin osiin.

#### **4.1.2 Fyysiset erot**

Usein ajatellaan, että suurin osa ihmisistä pystyy fyysisesti käyttämään suurinta osaa nykyisistä tietotekniikan sovelluksista. On kuitenkin paljon erityisryhmiä, joille monien sovellusten käyttö on hankalaa. Mieleen tulevat eri asteiset näkövammaiset, kuulovammaiset, liikuntavammaiset, käsivammaiset, pitkäaikaissairaat ja vanhukset. Ergonomisissa sovelluksissa tyypillisiä erityisryhmiä ovat mm. erittäin pitkät ja lyhyet ihmiset sekä vasenkätiset ja värisokeat.

#### **4.1.3 Psyykkiset erot**

Ihmiset poikkeavat huomattavasti toisistaan myös psyykkisen toimintakykynsä perusteella. Hahmottaminen voi olla esimerkiksi erilaista samoin päättelykyky ja oppimistyyli (Nielsen 1993). Toiset esimerkiksi kohtaavat uuden käyttöliittymän selailemalla ja kokeilemalla, toiset taas analyyttisesti etenemällä (Sinkkonen ym. 2002). Oppimisteorioissa puhutaan myös visuaalis-, auditiivis- ja kinesteettispainotteisista ihmisistä, mikä tarkoittaa käytännössä sitä, että toiset ihmiset oppivat parhaiten näkemällä, toiset kuuntelemalla ja toiset tekemällä eli aistikanava vaikuttaa oppimistuloksiin.

Myös psyykkisen toiminnan mukaan voidaan jaotella useita erityisryhmiä, esimerkiksi kehitysvammaiset, CP-vammaiset ja vanhukset. Edellä mainittujen lisäksi on monia lievempiä, ehkä vähemmän tunnettuja, havainnointiin ja oppimiseen vaikuttavia häiriöitä ja kognitiivisia toiminnan häiriöitä, jolle on ominaista kyvyttömyys

tunnistaa tuttuja esineitä ja asioita huolimatta aistien normaalisesta toiminnasta.)

- dysfasia (Aivoperäinen puuehäiriö, joka ilmenee puheen tuottamisen vaikeutena ja/tai puheen ymmärtämisen vaikeutena.)
- dysgrafia (Aivoperäinen kirjoittamishäiriö.)

- dysleksia (Aivoperäinen lukemisvaikeus. Käytetään yleisnimityksenä lukemisen ja kirjoittamisen häiriölle.)
- lukihäiriö, lukivaikeus (Näillä määritelmillä tarkoitetaan synnynäistä eri asteista vaikeutta oppia lukemaan ja kirjoittamaan. Lukihäiriö tai lukivaikeus voi esiintyä erillisenä tai yhdessä muiden oppimishäiriöiden kanssa.)
- MBD-oireyhtymä (Minimal Brain Dysfunction, aivotoiminnan häiriö, joka usein ilmenee tarkkaavuuden, motoriikan ja hahmottamisen häiriöinä, impulsiivisuutena ja oppimishäiriöinä.)

(HERO 2003)

#### **4.1.4 Kulttuuriset erot**

Kulttuurisia eroja on helppo nähdä eri kansallisuuksien välillä kielestä arkielämän tapoihin. Kulttuurisia eroja varten ei tarvita suurta maantieteellistä etäisyyttä, vaan esimerkiksi kaupunginosat, sosiaaliset kerrokset, sukupuoli tai vaikka harrastukset voivat muodostaa alakulttuureita. Kulttuuri ei myöskään ole paikkaan sidottua, vaan alakulttuuri voi elää internetin välityksellä ympäri maailmaa. (Sinkkonen ym. 2002)

Kulttuuri vaikuttaa monella tavalla käyttäjän kokemukseen tuomalla erilaisia konventioita, jotka vaikuttavat ihmisen ajatteluun. Kulttuuri voi esimerkiksi määritellä tiedon esitysmuotoja, ohjata havainnointia (esim. tekstin lukusuunta eri kulttuureissa) ja vaikuttaa niiden tulkintaan (esim. värien merkitys) (Sinkkonen ym. 2002). Ikkunoiden käyttöjärjestelmien helppokäyttöisyydestä osa tulee nimenomaan konventioiden kautta. Käyttöliittymäelementit toimivat samalla tavalla eri sovelluksissa ja samat tai samankaltaiset toiminnot löytyvät samoista paikoista eri sovelluksista.

#### **4.1.5 Kokemuksen merkitys**

Kokemus vaikuttaa merkittävästi siihen, kuinka käyttäjä käyttää tai pystyy käyttämään jotain sovellusta. Kokemusta voidaan tarkastella

esimerkiksi CSE (Cognitive Systems Engineering) mallin mukaan, missä ihmisen osaaminen jaetaan kolmeen tasoon: tietotason, sääntötason ja taitotason. Tietotaso kuvaa aloittelevan käyttäjän toimintaa, jossa käyttäjän täytyy aktiivisesti prosessoida uusia eteen tulevia asioita ja se on yleensä suhteellisen hidasta. Sääntötasolla toiminta on jo opittu paremmin, mutta tiedot ovat hajanaisia yksittäisiä sääntöjä. Taitotasolla kokonaisuus on suhteellisen hyvin hahmottunut, eikä päättelyä tarvita paljoa, vaan toiminnot tapahtuvat automaattisesti jonkun signaalin herättäminä. (Sinkkonen ym. 2002) CSE:n kolmitasoinen jaottelu sopii myös usein käytettyyn aloittelija, keskitason osaaaja, asiantuntija (novice, intermediate, expert) jaotteluun.

Osaaminen ei ole välttämättä kaikenkattavaa, vaan käyttäjä saattaa esimerkiksi olla taitotasolla vain pienessä osassa laajaa järjestelmää ja toisaalta taitotasolla oleva käyttäjä siirtyy toimimaan alemmilla tasoilla, jos hän kohtaa normaalista poikkeavan tilanteen käytön aikana. Kokemus voi liikkua eri akseleilla. Yleiset tiedot maailmasta, yleiset tiedot tietokonejärjestelmistä, tiedot jostain erityisestä järjestelmästä tai sovelluksesta, sovellusalueen tuntemus, aikaisemmat kokemukset ja asenne vaikuttavat yhdessä siihen, kuinka käyttäjä oppii käyttämään jotain tiettyä sovellusta. Järjestelmäkokemus auttaa järjestelmän ja käyttöliittymäelementtien peruskäytössä, kun taas sovelluksen aihealueen tunteminen auttaa taustalla olevien mallien ja terminologian ymmärtämisessä (Nielsen 1993).

#### **4.1.6 Taustatieto ja mentaalimallit**

Laajemmin tarkasteltuna kokemus ja kulttuuri muodostavat suuren massan taustatietoa ihmisille. Tuo massa on jokaisella yksilöllä erilainen ja kun se vaikuttaa siihen, miten yksilö kohtaa tilanteen, voidaan ajatella, että jokainen kohtaaminen on erilainen.

Kognitiivisen psykologian piirissä puhutaan skeemoista ja mentaalimalleista, jotka tarkoittavat tietorakenteita, joihin kaikki, esimerkiksi kulttuurin ja kokemuksen mukanaan tuoma tieto, on aivoissa varastoitunut. Skeemat ja mentaalimallit ohjaavat ihmisen toimintaa ja havainnointia ja jalostuvat sitä mukaan, kun ihminen omaksuu lisää tietoa. (Neisser 1980, Sinkkonen ym. 2002).

#### **4.1.7 Muuttuvat ominaisuudet**

Ihmisten ominaisuuksia voi luokitella myös sillä perusteella, ovatko ne synnynnäisiä vai myöhemmin hankittuja. Monet ihmisen ominaisuudet muuttuvat ihmisen elämänkaaren aikana. Fyysiset perusominaisuudet kuten vartalon mittasuhteet ja elimistön toimintojen lähtötaso ovat synnynnäisiä. Osa aivoissa olevasta informaatiosta syntyy geneettisesti koodattuna ihmisen kehityksen yhteydessä. Suurin osa informaatiosta tulee aivoihin aistien ja kognitiivisten prosessien kautta ihmisen elämän aikana (Neisser 1980).

Synnynnäisten lähtökohtien pohjalta ihminen muokkautuu ympäristön ja omien toimiensa seurauksena. Fyysisiä ominaisuuksia voi treenata, lisätä voimaa, parantaa ketteryyttä ja harjoittaa sorminäppäryyttä. Aivot ovat mukautuva elin, mikä tarkoittaa käytännössä sitä, että aivoihin tallentuva tieto vaikuttaa uuden tiedon vastaanottamiseen. Kulttuuriset ominaisuudet ovat opittuja tietoisesti tai tiedostamatta.

Ominaisuudet saattavat myös heikentyä, jos niitä ei pidetä aktiivisesti yllä. Esimerkiksi jonkun pitkään käyttämättä olleen sovelluksen käyttö ei välttämättä onnistu enää samalla tasolla tauon jälkeen (Sutcliffe 2000).

#### **4.1.8 Motivaation merkitys**

Motivaatio on olennainen tekijä ihmisen toiminnassa. Käyttöliittymien kannalta sillä on merkitystä, kun mietitään kuinka paljon käyttäjä

on valmis näkemään vaivaa jonkun sovelluksen oppimiseksi ja käyttämiseksi.

Motivoitunut käyttäjä on yleensä valmis tekemään työtä opetellakseen ja käyttääkseen sovellusta. Esimerkiksi näkövammaiset selailevat WWW-sivuja puhesyntetisaattorin avulla, koska se on heille lähes ainoa järkevä mahdollisuus päästä käsiksi Internetin sisältöön. Ulkopuolisesta käyttö tuntuu lähinnä kiduttavalta touhulta, koska sivuja suunniteltaessa ei ole otettu erityisryhmien tarpeita huomioon. Samalla tavalla raha ja toimeentulo yleensä lisäävät motivaatiota eli jos käyttäjä joutuu käyttämään sovellusta selvitäkseen työtehtävistään, on hän yleensä valmis näkemään vaivaa käytön opetelemiseksi.

Vastaavasti moni vähemmän motivoitunut käyttäjä työllistää mieluummin sihteerinään kuin opettelee hankalan fax-laitteen käytön tai soittaa mieluummin asiakaspalveluun kuin käy etsimässä tietoa yrityksen sekavilta verkkosivuilta.

#### **4.2 Käyttötilanteet ovat erilaisia**

Käyttötilanteet ovat erilaisia. Ihmisen ja koneen vuorovaikutuksessa on mukana aina ympäristö. Ympäristöön liittyy fyysisiä ominaisuuksia, kuten lämpötila tai taustamelu, teknisiä ominaisuuksia kuten sähkö- ja tietoliikenneverkon olemassaolo sekä monimuotoisia sosiaalisia tekijöitä. Käyttäjät saattavat käyttää tuotetta yhdessä, samassa tilassa voi olla muita ihmisiä, jotka häiritsevät käyttöä tai häiriintyvät käytöstä. Käyttäjän tunne yksityisyyden säilymisestä järkkyy helposti.

Ympäristön ja käyttäjän suhteeseen liittyvät sekä konkreettiset ympäristöstä tulevat signaalit että käyttäjän sisäinen tila ja suhtautuminen ympäristöön (Suchman 1987). Näyttää siltä, että käyttöympäristön vaikutusta käyttötilanteisiin on tutkittu suhteellisen vähän,

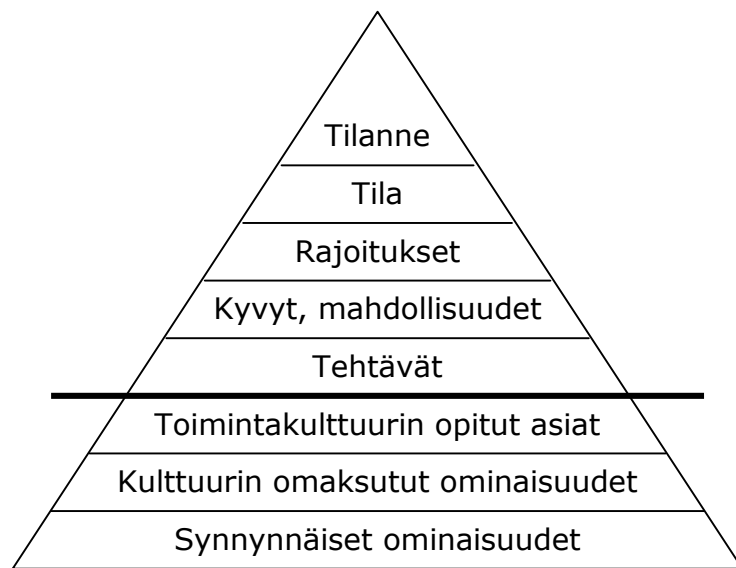
koska alan perusteoksissakin asiaa on käsitelty hyvin lyhyesti tai ei lainkaan.

Ilmeisesti syitä tutkimustiedon vähyyteen on ympäristön monimuotoisuudesta johtuva ilmiön vaikea mallintaminen ja tutkiminen, käyttötilanteiden keskittyminen toistaiseksi käytettävyysslaboratorion kaltaiseen toimistoympäristöön ja tietysti se, että tutkittavaa on riittänyt muutenkin. Ympäristön merkitys on tullut kuitenkin yhä tärkeämmäksi kannettavien tietokoneiden ja mobiilien päätelaitteiden tultua markkinoille sekä tietokoneiden sulautuessa tähän saakka epäilykkäisiin arkielämän esineisiin kuten autoihin ja kodinkoneisiin.

### **4.3 Moniulotteinen maali**

Edellä on kuvattu käyttäjän ja käyttötilanteiden monimuotoisuutta. Käyttöliittymäsuunnittelijan täytyisi pystyä liikkumaan tuossa moniulotteisessa maailmassa ja suunnittelemaan tuotteita, jotka tarjoavat käyttäjälle järkevän (hauskan/turvallisen/tehokkaan jne.) käyttökokemuksen.

Käyttökokemukseen liittyy valtava tietomäärä monelta eri tieteen alalta ja sen hallitsemiseen harva yksittäinen ihminen kykenee, kun se on vaikeaa monitieteellisille tutkijaryhmillekin. "Käytettävyyden psykologia" -kirjassa (Sinkkonen ym. 2002) esitellään kaavio (kuva 4.1), joka kuvaa tuotteen käyttämiseen liittyviä asioita. Vahvennetun viivan alapuolella olevat asiat ovat hitaammin muuttuvia ja niiden arviointiin voidaan käyttää aikaisempaa tutkimustietoa. Viivan yläpuolella olevat asiat täytyy käytännössä tutkia jokaisessa projektissa erikseen. Ongelmana on myös yhteisen teoriataustan puute. Eri osa-alueille on kehitetty erilaisia teorioita, joita voidaan hyödyntää, mutta niiden yhteensovittaminen saattaa olla hankalaa erilaisen lähestymistapojen takia.



Kuva 4.1. Toiminta- ja käyttöympäristöpyramidi (Sinkkonen ym. 2002)

Suuri ongelma on myös se, että suunnittelija ei välttämättä pysty kunnolla asettumaan käyttäjän rooliin. Siihen voi olla syynä yksinkertaisesti se, että ihmiset ovat erilaisia ja siinä mielessä myös suunnittelija on erilainen kun käyttäjä. Toisaalta tutustuttuaan alan tietoihin ja useisiin eri järjestelmiin, suunnittelija on voinut tulla sokeaksi niille asioille käyttäjäkokemuksessa, jotka ovat käyttäjälle tärkeitä, koska hänen omat mentaalimallinsa ja skeemansa asiasta ovat muuttuneet opiskelun myötä.

Donald Norman on esitellyt tilanteesta kolmikantaisen mallin, jonka osat ovat suunnittelumalli, käyttäjän malli ja järjestelmäkuva. Suunnittelumalli tarkoittaa suunnittelijan käsitteellistä mallia järjestelmän toivotusta toiminnasta ja käyttäjän malli tarkoittaa käyttäjän järjestelmästä muodostamaa käsitteellistä mallia. Suunnittelijan täytyy käytännössä kommunikoida haluamansa käsitteellinen malli järjestelmäkuvan (ulkomuodon, toiminnan, reagoititavan, käyttöohjeiden jne.) kautta käyttäjälle. (Norman 1989)



#### **4.4 Ratkaisutapoja**

Suunnittelijan ongelmaa asettua käyttäjän saappaisiin on pyritty ratkomaan valjastamalla insinööritaidot käyttöliittymäsuunnittelun palvelukseen. Käyttäjän toimia pyritään analysoimaan etukäteen ja toisaalta suunnitelmia testataan eri vaiheissa varsinaisten käyttäjien edustajilla, jotta nähdään ymmärtävätkö he käyttöliittymän samalla tavalla kuin suunnittelija on tarkoittanut tai miten helppoa tai vaivalloista tehtävästä suoriutuminen on. (Nielsen 1993)

Ihmisten erilaisuuden haasteeseen on pyritty tarttumaan eri tavoin. Yksi tapa on pienimmän yhteisen nimittäjän menetelmä, jossa pyritään tekemään käyttöliittymästä yksi kompromissi, joka soveltuu tarpeeksi hyvin kaikille käyttäjäryhmän edustajille. Usein ongelmaksi tulee se, ettei liittymä ole kenellekään paras mahdollinen.

Toinen tapa on suunnitella suurimmille aliryhmille omat käyttöliittymänsä. Esimerkiksi peruskäyttäjälle omansa ja ylläpitokäyttäjälle omansa.

Suunnittelija voi myös mahdollistaa saman asian tekemisen monella eri tavalla, jolloin käyttäjä voi valita itselleen sopivimman. Tätä on käytetty esimerkiksi Photoshop-kuvankäsittelyohjelmassa, jossa saman asian voi tehdä usealla eri työkalulla.

Yksi mahdollisuus on personointi, joka tarkoittaa sitä, että käyttäjä voi itse muokata käyttöliittymää vastaamaan omia mieltymyksiään. Asiantunteville käyttäjille personointi saattaa olla käyttökelpoinen työkalu. Jacob Nielsen on kuitenkin todennut, että käyttäjät eivät ole suunnittelijoita. Personointimahdollisuus monimutkaistaa käyttöliittymää, vaatii opettelua ja aikaa ja vaikeuttaa vertaistuen saamista, koska ihmisillä on erinäköiset käyttöliittymät. Käyttäjät eivät myöskään tiedä välttämättä, mikä on heille itselleen paras vaihtoehto. (Nielsen 1993)

Personointi voidaan suorittaa myös automaattisesti järjestelmän toimesta, jolloin puhutaan usein adaptiivisuudesta. Adaptiivisuuden haaste on käyttäjän hallinnan tunteen säilyttäminen tai käyttäjän suostuttelemine luovuttamaan hallintaa järjestelmälle ilman, että hän ahdistuu liikaa.

Multimodaalisuus on uusi mahdollisuus käyttöliittymäsuunnittelijoille. Usean aistikanavan käyttäminen mahdollistaa erilaisten käyttäjien ja käyttötilanteiden huomioimisen entistä paremmin.

#### **4.5 Suunnittelijan työkalut**

Käyttöliittymäsuunnittelijalla on arsenaali työkaluja, joiden avulla hän voi yrittää tehdä parhaat mahdolliset valinnat edellä kuvatussa moniulotteisessa suunnitteluavaruudessa.

Kognitiivinen psykologia tarjoaa teorioita ja käytännön tutkimustietoa siitä, kuinka ihmiset havaitsevat ja toimivat. Suuri osa HCI-alan kirjallisuudesta perustuu kognitiiviseen psykologiaan. Suunnittelija voi käyttää psykologian tietoja perussuunnitelman luontiin ja eteen tulevien suunnitteluongelmien ratkaisemiseen.

Suunnittelija voi myös tutkia käyttäjän toimintaa kokemuksellisesti. Jos tuote tulee valmiiseen tarpeeseen voi suunnittelija tarkkailla ja haastatella eri tavoin tuotteen tulevia käyttäjiä. Mikäli kysymyksessä on täysin uusi tuote, jolle ei ole vielä valmista käyttäjäkuntaa, voi suunnittelija yrittää eläytyä tai pyytää muita ihmisiä eläytymään käyttäjän asemaan ja rakentaa erilaisia skenaarioita tuotteen käyttötilanteista.

Jotta käyttöliittymän toimintaa voitaisiin kuvata käyttäjälle tarvitaan yleensä prototyyppi. Prototyyppejä voi olla monen eri tasoisia riippuen suunnitteluvaiheesta. Graafisten käyttöliittymien yhteydessä

käytetään suunnittelun alkuvaiheessa usein ns. paperiprotoja, jotka ovat piirroksia tulevasta käyttöliittymästä. Käyttäjät voivat sitten kertoa, miten he toimisivat, jos käyttöliittymä olisi oikea. Puhekäyttöliittymien ja monien muiden modaliteettien kohdalla ei voida käyttää paperiprotoja. Tällaisissa tapauksissa voidaan käyttää ns. Wizard-of-Oz testausta, jossa ihminen matkii järjestelmän toimintaa. Wizardin roolissa oleva ihminen tunnistaa esimerkiksi käyttäjän puheet tai eleet ja niiden perusteella antaa toiminnon palautteen tai ohjaa järjestelmän antamaan palautteen käyttäjälle. Wizardin rooli on multimodaalisten käyttöliittymien yhteydessä vaativampi, koska koordinoitavien asioiden määrä kasvaa (Hemsen 2002). Käytännössä kannattaa harkita useamman wizardin käyttöä.

Myöhemmissä vaiheissa prototyypit voivat olla teknisen alustan päälle rakennettuja ja oikeasti vuorovaikutteista toiminnallisuutta sisältäviä versioita. Tunnistuspohjaisissa vuorovaikutustavoissa (puheentunnistus, eleiden tunnistus jne.) testaus tuotantotekniikalla on erityisen tärkeää, koska käytetty tekniikka vaikuttaa ratkaisevasti tunnistustuloksiin.

Jotta voidaan varmistua siitä, että suunnittelija on onnistunut tehtävässään täytyy tuote testata tulevien käyttäjien edustajilla. Tavallisissa graafisissa sovelluksissa yleinen testaustapa on ääneen ajattelu, jossa testikäyttäjä kertoo ääneen, mitä hän tekee, miksi hän tekee niin ja miksi joku tehtävä ei tahdo onnistua. Näin suunnittelija pääsee käsiksi käyttäjien erilaisiin käyttötilanteeseen liittyviin sisäisiin malleihin. Puhekäyttöliittymän yhteydessä ääneen ajattelu ei luonnollisesti toimi, koska puhuminen kesken vuorovaikutustilanteen sotkisi tehtävän kulun ja toisaalta sotkisi mahdollisesti käytössä olevan puheentunnistusjärjestelmän toiminnan. Eräs ratkaisu on tallentaa käyttötilanne videolle ja toistaa se heti tehtävän jälkeen testihenkilölle, joka voi sitten kommentoida tuntemuksiaan käyttötilanteen aikana (Vaittinen 2003).

Kokemuksen kasvaessa huomataan, miten teoriat toimivat käytännössä ja mitä ilmiöitä käyttötilanteisiin liittyy. Hyviksi havaitut suunnitteluperiaatteet voidaan koota suunnitteluohjeistoiksi, joita voidaan käyttää myöhemmin hyväksi. Suunnitteluohjeistot voivat olla eri tasoisia, yleisiä kaikkiin käyttöliittymiin päteviä, johonkin tiettyyn käyttöliittymätyyppiin kuten puhekäyttöliittymiin liittyviä ja yritys- tai tuotekohtaisia (Nielsen 1993). Sekä graafisille, että puhekäyttöliittymille löytyy useita suunnitteluohjeistoja. Multimodaalisille käyttöliittymille kunnollisia suunnitteluohjeistoja ei vielä ole. Kirjallisuudesta löytyy jotain suppeita ja hajanaisia ohjeita. Multimodaalisten käyttöliittymien suunnittelussa on mahdollista soveltaa yksittäisten modaliteettien omia suunnitteluohjeistoja, mutta modaliteettien yhdistämistä koskevat päätökset joudutaan usein testaamaan käytännössä. (Vaittinen 2003)

Edellä mainittujen työkalujen ja -menetelmien käyttämiseen hyvässä järjestyksessä tarvitaan jonkinlaista prosessia. Jakob Nielsenin (1993) tunnetuksi tekemät käytettävyyden insinööritaidon (Usability Engineering) periaatteet lienevät tunnetuin kuvaus käyttöliittymäprosessin kulusta. Muitakin vastaavia on, kuten kansainvälisen standardointiorganisaation (ISO) kehittämä standarditapa (ISO 13407 1999) . Yhteistä näille prosesseille on iteratiivisuus eli suunnittelua tarkennetaan useamman suunnittelukierroksen aikana käyttäjätesteistä saadun palautteen perusteelle. Iteratiivisen suunnittelun helpottamiseksi olisi hyvä, jos käytössä olisi työkaluja, joilla käyttöliittymäprototyypin muokkaus kävisi nopeasti (RAD, Rapid Application Development tools). Multimodaalisille sovelluksille tällaisia työkaluja ei vielä ole olemassa, mutta tiettyjä modaliteettiyhdistelmiä varten sellaisia ollaan kehittämässä (Hemsen 2002).

#### **4.6 Rakennuspalikat**

Käyttöliittymäsuunnittelijalla on käytössään valikoima rakennuspalikoita. Usein rakennuspalikat tulevat annettuina teknisen alustan puolelta (vrt. Windows tai jossain määrin WWW), joskus suunnittelija pääsee tai joutuu tekemään ne itse (vrt. Macromedia Flash). Valmiiksi annetut palikat helpottavat suunnittelijan työtä vähentämällä työmäärää ja rajaamalla suunnitteluavaruutta. Ne myös edistävät yhtenäisen käyttökokemuksen muodostumista saman alustan päälle tehtyjen sovellusten kesken. Haittapuoli on se, että ne rajoittavat joissain tilanteissa optimaalisten käyttöliittymäratkaisujen tekemistä.

Graafisissa käyttöliittymissä peruspalikoita ovat esimerkiksi ikkunat, napit, valikot, ikonit, kohdistimet, linkit, lomakkeet, asettelu, teksti, typografia, värit, kuvat, video ja animaatio.

Myös puhekäyttöliittymiin on olemassa jonkun verran järjestelmätoimittajakohtaisia peruspalikoita (Balentine & Morgan 1999), mutta aika paljon asioita joudutaan käytännössä tekemään sovelluskohtaisesti. Puhekäyttöliittymien peruselementtejä ovat esimerkiksi sanavalinnat, lauserakenteet, tauot, äänensävy, intonaatio, valikkohierarkiat, "barge-in" (ks. kappale 5.8.1), avainsanat, keskustelun aloitteen määrääminen ja muut äänet (earcons, sound icons).

Käyttöliittymäsuunnittelussa käytetään usein metaforia, joiden avulla liittymän toimintaa yritetään avata käyttäjälle. Graafisissa käyttöliittymissä tyypillisiä metaforia ovat työpöytä, ikkuna, lomake ja painike. Puhekäyttöliittymissä perusmetafora on yleensä keskustelu. Lisäksi voidaan ajatella, että erilaisilla hahmoilla voidaan vaikuttaa siihen, miten ihminen suhtautuu käyttöliittymään. Esimerkiksi, jos puhekäyttöliittymä esittää selkeästi tyhmää konetta tai avaruusolioita ihmisen sijaan, saattaa käyttäjä kokea käyttötilanteen erilaisena ja esimerkiksi hyväksyä enemmän järjestelmästä johtuvia virheitä.

Graafisissa käyttöliittymissä voidaan käyttää moodeja eli käyttötiloja, joissa esimerkiksi näppäinten toiminnot muuttuvat riippuen siitä, missä tilassa ollaan. Puhekäyttöliittymissä modeeja voidaan simuloida käyttämällä eri hahmoja samassa sovelluksessa. Hahmot voidaan erottaa toisistaan esimerkiksi äänensävyn, puhutavan ja sukupuolen perusteella. Puhesovelluksessa voi olla esimerkiksi naisääni normaaleissa toiminnoissa ja miesääni ohjetoiminnossa.

Antropomorfisessa eli inhimillistävässä suunnittelussa on hyvät ja huonot puolensa, mutta käytännössä se on niin moniulotteinen asia, että se on melkein oma tieteen/taiteenlajinsa (Balentine & Morgan 1999, Boyce 2000) (ks. myös 7.3).

#### **4.7 Syöte- ja tulostuslaitteet**

Käytössä olevat syöttö- ja tulostuslaitteet vaikuttavat monella tavalla multimodaaliseen kokemukseen.

Laitteisiin liittyvät affordanssit vaikuttavat käyttökokemukseen (ks. kappale 5.1). Käyttäjän kannalta on esimerkiksi eri asia näkeekö hän mikrofonin, johon voi puhua vai pitääkö hänen vain tietää, että sovellukselle on mahdollista puhua, jos mikrofoni on piilossa.

Laitteiden valinnalla voidaan vaikuttaa myös yksityisyyden suojaan. Kuulokkeiden avulla äänipalautteet voidaan kohdistaa ainoastaan yhdelle käyttäjälle ja pieni kannettava näyttö (esim. PDA) on helppo pitää sivullisilta piilossa.

Teknisestä näkökulmasta laitteiden valinnalla on myös merkitystä. Esimerkiksi sankamikrofoni mahdollistaa sen, että äänenvoimakkuus pysyy järjestelmän kannalta tasaisena, koska mikrofonin etäisyys suusta pysyy vakiona.

Seuraavassa on esitelty tyypillisimmät graafiset näyttölaitteet, puhekäyttöliittymän laitteet ja osoituslaitteet:

Graafiset näytöt:

- Kuvaputkinäyttö. Sopii toimistokäyttöön ja informaatiokioskeihin. Mahdollistaa tarkan kuvan ja suuren tietomäärän esittämisen.
- Litteä näyttö (esim. LCD). Sopii toimistokäyttöön ja informaatiokioskeihin. Mahdollistaa tarkan kuvan ja suuren tietomäärän esittämisen. Sopii myös liikkuvaan käyttöön (vrt. Tablet PC tai PDA) ja ajoneuvoihin.
- Dataprojektori. Sopii mediatiloihin ja useamman yhtäaikaisen käyttäjän sovelluksiin. Kallis.
- Head Up Display (HUD), käyttäjän päähän kiinnitettävä näyttö eli silmälappuvideo. Sopii kannettaviin sovelluksiin. Jättää kädet vapaaksi. Rajoitettu resoluutio ja kuvanlaatu.

Puhekäyttöliittymän laitteet:

- Pöytämikrofoni. Sopii toimistokäyttöön ja informaatiokioskeihin. Vaatii jonkinasteista puheen suuntaamista.
- Sankamikrofoni tai nappimikrofoni. Sopii toimistokäytön lisäksi liikkuvaan käyttöön. Puhetta ei tarvitse erikseen suunnata, varsinkaan sankamikrofonin kanssa.
- Mikrofonimatriisi (Microphone array). Joukko mikrofoneja, joiden avulla "mikrofonin" suuntausta voidaan muuttaa ohjelmallisesti. Mahdollistaa esimerkiksi mikrofonin suuntauksen liikkuvaa käyttäjää kohti.
- Kaiuttimet. Sopivat toimistokäyttöön, informaatiokioskeihin sekä ajoneuvoihin. Mahdollistaa useamman käyttäjän sovellukset.
- Kuulokkeet. Sopivat toimistokäyttöön ja liikkuvaan käyttöön. Mahdollistavat yksityisyyden säilyttämisen.

Osoituslaitteet:

- Hiiri. Edullinen ja vakiovarusteena nykyisissä pöytätietokoneissa. Vaatii pöytätilaa käyttöä varten.
- Ohjauspallo (Trackball). Väärinpäin käännetty hiiri. Ei vaadi pöytätilaa. Motorisesti erilainen kuin hiiri, koska kohdistinta ohjailaan peukalolla, ei koko kädellä.
- Datahanska. Mahdollistaa sormien asennon tutkimisen ja sitä kautta eleiden tunnistamisen. Joissain tapauksissa mahdollistaa osoittamisen tilassa. Kallis.
- Kosketusnäyttö. Kulkee näytön mukana. Mahdollistaa helpon suoraosoituksen näytöltä.
- Liikesensorit. Mahdollistavat osoittamisen tilassa.
- Silmäohjaus. Mahdollistaa osoittamisen kohdistamalla katseen osoitettavaan kohtaan. Vaatii harjoittelua käyttäjältä. Pitkään käytettynä rasittava. Kallis.

#### **4.8 Rahamaailman rajoitukset**

Kun tehdään tutkimus- tai taideprojekteja, voidaan useimmiten keskittyä johonkin ydinasiaan, jota varten kyseinen projekti on perustettu. Kun kaupallinen yritys tekee tuotteita myyntiin, tulee kuvioihin mukaan yleensä useampia pelureita ja suunnittelutyössä täytyy ottaa huomioon monia asioita ydinasian ulkopuolelta. Aina kun tehdään jotain, tarvitaan taloushallinnon kielellä resursseja eli käytännössä aikaa, rahaa, välineitä ja ennen kaikkea ihmisiä.

##### **4.8.1 Monimutkaisuuden seuraukset**

Modaliteettien mukana käyttöliittymän ja järjestelmän monimutkaisuus kasvaa. Monimutkaisuuden takia multimodaalisten käyttöliittymien aikaansaamiseen tarvitaan enemmän resursseja kuin perinteisten käyttöliittymien suunnitteluun ja toteuttamiseen. (Oviatt ym. 2000c)



Yhdenkin modaliteetin käyttöliittymän suunnitteluun tarvitaan asiantuntijoiden panosta ja eri modaliteeteilla on omat erityispiirteensä, joten ne vaativat usein oman asiantuntijansa. Lisäksi tarvitaan joku, joka ymmärtää, miten eri modaliteetit voidaan yhdistää, joten voidaan sanoa, että multimodaalisen käyttöliittymän suunnitteluun tarvitaan enemmän asiantuntijuutta kuin perinteisten käyttöliittymien suunnitteluun. Välttämättä ei tarvita enempää ihmisiä, yksikin ihminen voi omata laaja-alaisesti asiantuntijuutta, tässä tapauksessa usean eri modaliteetin osalta. Käytännössä kuitenkin asiantuntijoiden määrä on kääntäen verrannollinen asiantuntisuuden määrään eli mitä laaja-alaisempi asiantuntija, sitä vähemmän heitä on markkinoilla ja sitä enemmän he yleensä maksavat, alasta riippumatta.

Suunnittelu-aikaa tarvitaan myös enemmän, koska monimutkaisemmassa järjestelmässä on enemmän yksityiskohtia, joihin täytyy ottaa kantaa. Reaalista suunnittelu-aikaa voidaan yrittää lyhentää lisäämällä suunnittelijoiden määrää, mutta tässäkin tapauksessa pätee sama asia kuin edellisessä kohdassa asiantuntijoiden saatavuudesta ja heidän rahallisesta arvostaan. Lisäksi vastaan tulee vanha totuus sopasta ja useammasta keittäjästä eli mitä enemmän ihmisiä tarvitaan, sitä enemmän organisaatioon ja viestintään täytyy panostaa.

Vielä tällä hetkellä multimodaalisten järjestelmien toteuttamiseen tarvittava tekniikka on enemmän tai vähemmän kokeiluasteella. Valmiita käytännössä hyväksi hiottuja työkaluja ja ohjelmointiympäristöjä ei ole saatavilla samalla tavalla kuin esimerkiksi Windows-ohjelmien tai WWW-sivujen tekemistä varten. Teknisellä puolella tarvitaan siis samalla tavalla eri alojen asiantuntijoita kuin käyttäjäkokemuksen suunnittelussa.

Koska tekninen infrastruktuuri ei ole vakiintunut eikä tietämystä ole muutenkaan kertynyt vielä paljoa käytännön kautta, vaatii suunnittelu kaikilta mukanaolijoilta valistuneisuutta ja kommunikointikykyä.

Lisäksi jo olemassa oleva tekniikka on nykypäivänä monien modalityteettien osalta suhteellisen kallista. Hyvät puheentunnistimet, puhesyntetisaattorit, katseen- tai eleentunnistimet eivät ole edullisia. Hinnat tulevat todennäköisesti tippumaan, jos tekniikoille keksitään ja kehitetään sellaisia sovelluksia, joista kuluttajat ovat valmiita maksamaan.

Kustannusten lisäksi projektin rahoittajat haluavat yleensä tietää mahdolliset riskit. Multimodaalisten käyttöliittymien monimutkaisuus nostaa käytännössä epäonnistumisen riskiä. Mitä enemmän modalityteetteja, sitä suurempi riski.

#### **4.8.2 Markkinoinnin haasteet**

Jos tehdään kuluttajatuotteita, kuuluu kuvioihin yleensä markkinointiosasto. Tuote yritetään yleensä sovittaa yrityksen tai tuoteperheen imagoon tai brändiin.

Aika monella yrityksellä on nykypäivänä graafinen ohjeisto tietokonesovelluksia ja verkkosivuja varten, mutta käytännössä aika harvalla on esimerkiksi ohjeisto puhesovellusten tyyliin määrittelemiseksi. Puhesovelluksen tyyliin liittyy muun muassa puheääni, -nopeus, puhujan sukupuoli ja puhuttelutapa. Lisäksi tekniikka saattaa asettaa rajoituksia niin, ettei esimerkiksi puhesyntetisaattoria käytettäessä voida saada aikaan täysin halutunlaista ääntä.

Toinen mielenkiintoinen haaste tulee eteen, kun yritetään kommunikoida samaa brändiä usean eri modalityteetin kautta. Ajatusta voisi lähestyä arkikokemuksen kautta. Monille on varmaan muodostunut sarjakuvia lukiessa tietty mielikuva sarjakuvahahmojen äänestä.

Kun myöhemmin näkee saman sarjakuvan pohjalta tehdyn animaation elokuvateatterissa saattaa hämmästys olla melkoinen, kun hahmot kuulostavatkin aivan erilaisilta kuin oli aikaisemmin kuvitellut.

#### **4.8.3 Multimodaalisuuden mahdollisuudet**

Edellä kuvailtiin multimodaalisuuden vaikutusta suunnittelukustannuksiin (ja hieman tuotantokustannuksiinkin). Suurempien kustannusten ja riskien takia multimodaalisuuden täytyy käytännössä tuoda merkittäviä etuja niin, että käyttäjät ovat valmiita maksamaan tuotteesta kalliimman hinnan.

Suurin mahdollisuus on luonnollisesti siinä, että multimodaalisuuden avulla voidaan toteuttaa sellaisia sovelluksia, joita ei voida muuten toteuttaa tai sellaisia sovelluksia, joiden toteuttaminen yhtä modaliitteita käyttäen olisi vaikeaa. Esimerkiksi carMMI:n kaltaista auto-tietokonesovellusta, jossa käyttäjä voi ohjata karttoja esittävää navigointisovellusta ilman käsiä, on aika mahdoton toteuttaa. Pelkällä puheella toimiva navigointisovellus on mahdollista tehdä, mutta käytännössä esimerkiksi kaupunkitilanteessa tulee helposti ryhmittymis- ja risteystilanteita, joissa oikean haaran opastus pelkällä puheella on hankalaa.

Yrityksen imagon kannalta multimodaalisuutta voidaan käyttää merkkinä edistyksellisyydestä, hi-tec osaamisesta ja panostuksesta sekä tulevaisuuteen että käyttäjiin. Kalleimpien autojen varustuksiin kuuluu multimodaalisia navigointijärjestelmiä ja esimerkiksi Nokia tutkii multimodaalisten käyttöliittymien mahdollisuuksia matkapuhelimissa (Talouselämä 11/2003).

Multimodaalisuutta voidaan myös käyttää luomaan elämyksiä, joita tavallisella WIMP-käyttöliittymällä ei voida saavuttaa. Voi helposti kuvitella, että esimerkiksi peliteollisuudelle kokonaisvaltaisen käyt-

täjäkokemuksen saavuttaminen on yksi tärkeimmistä tutkimuskoh-  
teista. Pelit ovat myös siitä otollisia kohteita, ettei niiden suunnitte-  
lua rajoita reaali maailman vaatimukset samalla tavalla kuin monien  
hyötysovellusten kohdalla, vaan peliä suunniteltaessa voidaan tek-  
niikan rajoitukset ottaa optimaalisesti huomioon.

## 5. Modaliteettien ominaisuudet

Modaliteettien ominaisuudet vaikuttavat siihen, minkälaista tietoa ja missä muodossa ihminen pystyy välittämään ja vastaanottamaan niiden kautta. Visuaalinen ja auditiivinen modaliteetti eroavat monessa suhteessa toisistaan. Seuraavassa on pohdittu visuaalisen ja auditiivisen modaliteettien eroja graafisen ja puhekäyttöliittymän kautta.

### 5.1 *Affordanssit, rajoitukset ja konventiot*

Havaintopsykologi J. J. Gibson kehittämä ja Donald Normanin (1989, 1993, 1999) HCI-keskusteluun tuoma affordanssin käsite kuvaa mahdollisuuksia, joita käyttöliittymä tarjoaa. Mahdollisuudet liittyvät malleihin, joita ihmisillä on maailmasta. Norman on itse myöhemmin jakanut affordanssit kahteen luokkaan aitoihin (real affordances) ja havaittuihin (perceived affordances). Aito affordanssi tarkoittaa jotain fyysistä mahdollisuutta (ovenkahva mahdollistaa tarttumisen, näyttö mahdollistaa koskettamisen, kenttä weblomakkeessa mahdollistaa kirjoittamisen). Havaittu affordanssi tarkoittaa jotain toimintoa (oven avaus, näppäimen painaminen näytöllä), jonka käyttäjä havaitsee tuotteessa. Käyttäjää voi ohjata havaitsemaan affordanssin tarjoamalla vihjeitä (feedback of affordance, esim. ovenkahva, näppäimen graafinen kuva).

Lisäksi Norman (1989, 1999) esittelee rajoituksen ja konvention käsitteet. Rajoitukset hän jakaa fyysisiin, semanttisiin, kulttuurisiin ja loogisiin. Konventiot ovat ajan kuluessa muovautuneita kulttuurisia rajoituksia.

Graafisen käyttöliittymän tärkeimmät aidot affordanssit ovat näytön katsominen ja koskeminen, jos se on kosketusetäisyydellä sekä hiiren liikuttelu ja näppäinten painaminen. Aistitut affordanssit riippuvat sovelluksesta.

Fyysisiä rajoituksia puolestaan ovat esimerkiksi mahdollisen näppäimistön aiheuttamat rajoitukset tekstin syötössä ja kohdistimen liikkumisen rajoittaminen näytön sisäpuolelle. Loogiset ja kulttuuriset rajoitukset voivat olla moninaiset. WIMP-käyttöliittymissä ne liittyvät ikkunoiden hallintaan, valikkojen toimintaan jne.

Puhekäyttöliittymän suhteen affordanssi on mielenkiintoinen käsite. Periaatteessa käyttäjä voi puhua aivan vapaasti mitä haluaa, milloin haluaa ja missä haluaa. Voidaan ehkä ajatella, että jos järjestelmä tekee aloitteen vuorovaikutustilanteessa, toimii se ensimmäinen puheenvuoro vihjeenä puhumisen affordanssista käyttäjälle.

Käytännössä rajoituksilla onkin suurempi merkitys puhekäyttöliittymissä. Esimerkiksi mikrofoni rajoittaa puhumisen tiettyyn paikkaan ja ihmisten välisessä vuorovaikutuksessa opittuja konventiota voidaan käyttää ohjaamaan käyttäjää puhekäyttöliittymädialogin aikana.

## **5.2 Tiedon ajallinen kesto**

Graafinen ja puhekäyttöliittymä ovat erilaisia tiedon esityksen ajallisen keston suhteen. Lukuun ottamatta tiettyjä sovelluksia, kuten pelejä, joissa tiedon ajallista kestoja rajoitetaan tarkoituksella, pysyy tieto graafisen käyttöliittymän ruudulla käytännössä niin kauan kuin käyttäjälle on tarpeen. Puhekäyttöliittymässä tieto häviää kuuluvista saman tien, kun dialogi etenee. Käyttäjän kannalta se jää elämään ainoastaan hänen aivojensa tietojenkäsittelyprosesseihin. Graafinen käyttöliittymä on siis joustavampi tiedon ajallisen esitystavan suhteen.

Eryteisesti puhekäyttöliittymien puolella tulee ihmisen työmuistin pienuus esiin tiedon lyhyen ajallisen keston takia. Työmuisti pystyy käsittelemään yhtäaikaan 3-5 mieltämysyksikköä, jotka ovat riippuvai-

sia yksilön aivoihin aiemmin muotoutuneista tietorakenteista. Käyttäjän ei voida olettaa muistavan suuria tietokokonaisuuksia, jotka on ilmaistu kerran puheena. Dialogissa ei yhden vuoron aikana kannata esittää liikaa erillisiä muistettavia asioita. Esimerkiksi, jos puhevalikossa on useampia kohtia, toimenpiteen ehto kannattaa esittää ensin ja komentosana vasta sen jälkeen, jottei komentosana unohtuisi ehtoa kuunnellessa (Sinkkonen 2002, Weinschenk & Barker 2000: 250).

Puheena esitetyn täsmällisen tiedon muistaminen on niin ikään vaikeaa eli käyttäjän ei voida olettaa muistavan esimerkiksi tarkkoja lukuarvoja tai puhelinnumeroita. Nykyaikaisissa numerotiedustelupalveluissa asia on ratkaistu puheluiden yhdistämispalvelulla ja lähettämällä numero tekstiviestinä.

### **5.3 Rinnakkainen vs. sarjamuotoinen tieto**

Visuaalinen modaliteetti mahdollistaa tiedon rinnakkaisen esittämisen. Esimerkiksi palvelun rakenne voidaan näyttää puumaisena esityksenä ruudun sivussa, samalla kun palvelun sisältöä esitetään ruudun keskellä. Puumainen rakenne voi lisäksi toimia vuorovaikutteisena elementtinä niin, että sen avulla voidaan siirtyä palvelun eri osioihin.

Rinnakkaisuus voi toimia graafisella näytöllä monella eri tasolla. Esimerkiksi Windows-käyttöliittymässä, jossa käytetään ikkuna-metaforaa perusrakenteena, voi yhden ikkunan sisällä olla useita rinnakkaisia osioita, ikkunat voivat olla sisäkkäin toisten ikkunoiden ali-ikkunoina tai rinnakkain toisistaan irrallisina. Lisäksi eri tarkoitusta varten olevat ohjelmat voivat olla auki yhtä aikaa, ja niiden käyttöliittymät voivat olla rinnakkain näkyvissä, jolloin käyttäjällä on mahdollisuus käsitellä eri ohjelmien tuottamaa informaatiota yhtä aikaa.

Mahdollisuus esittää tietoa rinnakkain mahdollistaa esimerkiksi asioiden suhteiden visuaalisen esittämisen. Pylväsdiagrammilla voidaan esittää osakekurssien vaihtelua ajan suhteen tai puurakenteella www-sivujen hierarkkisia suhteita toisiinsa navigointikartassa.

Auditiivisen modaliteetin kautta tieto esitetään sarjamuodossa eli peräkkäin. Ihmisen on lähes mahdotonta ymmärtää kahden yhtäaikaan puhutun tekstin sisältöä (Neisser 1981). Sarjamuotoisesta esitysmuodosta aiheutuu muutamia ilmiötä. Tilallisia suhteita on vaikea ilmaista, jos ne eivät ole valmiiksi kielen käsitteitä. Esimerkiksi karttaa tai palvelun hierarkiaa ei ole helppo esittää puhekäyttöliittymässä. Samoin monimutkaisia asioita, esimerkiksi laitteen rakenne, on myös vaikea esittää puheena, jos ne eivät ole valmiiksi käsitteitä. Vaikka organisaation suhteet voidaan ilmaista kielellisesti ("Sirpa on Riston pomo"), on vähänkin suuremman yrityksen koko organisaattorakenne käytännössä erittäin vaikeaa omaksua yhtenä pitkänä puheesityksenä.

Sarjamuodossa on myös etunsa. Tieto voidaan esittää käyttäjälle juuri silloin, kun hän sitä tarvitsee. Käyttäjän ei tarvitse etsiä tietoa näytöltä, vaan se voidaan esittää juuri suunnitellussa kohdassa dialogin mukana.

Kuuloaistilla on myöskin rinnakkainen ulottuvuus siinä mielessä, että se vastaanottaa tietoa myös kuuntelijan tarkkaavaisuuden kohteen ulkopuolelta. Ominaisuus tulee esille ns. "Cocktail party" -ilmiössä, jossa ihminen saattaa tunnistaa kovan puheensorinan keskeltä oman nimensä (Dix ym. 1998). Ihminen voi myös oppia hyvin rajallisessa määrin tarkkailemaan valikoidusti kahta puhevirtaa rinnakkain (Neisser 1981)



#### **5.4 Tunnistaminen vs. muistaminen**

Olennainen osa ihmisen tiedonkäsittelyä on aikaisemman tiedon muistiin palauttaminen. Muistiin palauttamisesta voidaan erottaa kaksi eri vaihetta: muistaminen ja tunnistaminen. Muistamisessa informaatio täytyy palauttaa mieleen tilanteessa olevia hakuavaimia käyttäen, kun tunnistamisessa täytyy vain muistaa, että mielessä oleva informaatio vastaa aikaisempaa muistijälkeä (Sinkkonen 2002).

Tunnistamisen ja muistamisen ero tulee käyttöliittymissä esille useilla tavoilla. Komentorivipohjaisissa käyttöliittymissä (esim. MSDOS, Unixit) käyttäjän täytyy muistaa komennot ja niiden käyttötavat ulkoa, koska komentokehote (esim. C:\>) ei anna mitään vinkkejä käytettävissä olevista komennosta. Tällainen käyttöliittymä voi olla nopea käyttää, mutta vaatii aika paljon opetteluja ja lisäksi hyvän hahmotuskyvyn, jotta pystyy ymmärtämään ja käyttämään esimerkiksi hierarkkista hakemistorakennetta kunnolla hyväkseen.

Graafisessa käyttöliittymässä käytettävissä olevat toiminnot voidaan tehdä näkyväksi. Esimerkiksi usein käytetyissä pudotusvalikoissa on rajattu lista komentoja, joista käyttäjä valitsee yhden. Riittää, että käyttäjä tunnistaa haluamansa komennon muiden joukosta, jolloin tunnistaminen tapahtuu komennon nimen perusteella. Toisaalta käyttäjä esimerkiksi tunnistaa linkin muun tekstin joukosta, koska se eroaa tavallisesta tekstistä alleviivauksen takia ja koska se on websivulla, joissa alleviivausta on käytetty perinteisesti linkin tunnistena.

Puhekäyttöliittymissä perustilanne on vähän samantapainen kuin komentorivipohjaisissa käyttöliittymissä. Järjestelmä kuuntelee käyttäjän syötteitä ja käyttäjän täytyy tietää, mitä sille voi sanoa. Koska taustalla oleva järjestelmä ei ymmärrä mitä tahansa, täytyy mahdollisuudet kommunikoida käyttäjälle jollain tavalla. Jos kysees-

sä on käyttäjän usein käyttämä järjestelmä ja käytetty sanasto on rajattu, on mahdollista opetella tarpeelliset komentosanat, jolloin käyttö on samantyyppistä kuin komentorivipohjaisessa käyttöliittymässä. Tällainen on esimerkiksi hammaslääkärin sovellus, jossa lääkäri luettelee hammastarkastuksen aikana diagnoosinsa tietokoneelle. Tietokone tulkitsee sanelun ja merkitsee tiedot hammaskarttaan eli tekee työn, jonka hammashoitaja tekee normaalisti. Sovellus toimii hyvin käytännössä, koska lääkäri joutuu joka tapauksessa opettelemaan käytetyn komentokielen eli diagnoosisanaston ja muistaa sen ulkoa, koska joutuu käyttämään sitä joka päivä.

Jos sanasto on laaja tai käyttäjä tarvitsee sovellusta harvoin, täytyy toiminnot ja komennot tehdä jollain tavalla näkyviksi käyttötilanteessa. Järjestelmän tuottaman dialogin täytyy jollain tavalla ohjata käyttäjää sanomaan oikeita sanoja tai lauseita. Yksinkertaisin tapa on luetella kaikki mahdolliset sanat, joista käyttäjä sitten valitsee yhden lausumalla sen (eksplisiittinen kehote). Tällä tavoin puheentunnistustehtävä on järjestelmän kannalta helpompi, mutta käyttäjän kannalta dialogista tulee usein kankeaa ja holhoavan tuntuista ("Jos haluat saada säätiedot, sano SÄÄ, jos haluat saada osakekurssit sano OSAKKEET"). Toinen mahdollisuus on muotoilla dialogi siten, että käyttäjää ohjataan sanomaan halutut asiat ilman, että kaikkia sanoja luetellaan käyttäjälle ("Haluatko kuulla säätiedot vai osakurssit?") (implisiittinen kehote). Tällöin dialogin muotoilu on erittäin olennaisessa osassa. Ei ole aivan yksinkertainen tehtävä ohjata ihmisiä sanomaan juuri haluttuja asioita. Mitä heterogeenisempi käyttäjäpohja on, sitä hankalammaksi tehtävä muuttuu. Dialogisuunnittelua on tutkittu puhe käyttöliittymien yhteydessä ja aiheesta on saatavilla yksityiskohtaisia suunnitteluoppaita, esimerkiksi Weinschenk ym. (2000) ja Balentine ym. (1999).

### **5.5 Suunnitelmallinen vs. tilannekohtainen toimintatapa**

Ihmisen toimintaa voidaan kuvata suunnitelmalliseksi tai tilannekohtaiseksi riippuen siitä, mikä toiminnan motivoi. Suunnitelmallinen, pohdiskeleva toiminta lähtee ihmisen sisäisestä tarpeesta ja on tietoista. Ihminen toimii pohdiskelevassa tilassa silloin, kun hän miettii asioita, luo ja tutustuu uusiin asioihin kuten esimerkiksi uuteen käyttöliittymään. Suunnitelmallinen toiminta on suhteellisen hidasta ja työlästä.

Tilannekohtainen, kokemuksellinen toiminta taas lähtee ulkoisesta ärsykkeestä. Kokemuksellisessa tilassa ihminen vastaanottaa suuren määrän ulkopuolista informaatiota eri aistikanavien kautta ja toiminta tapahtuu sisäisen tiedon pohjalta automaattisesti ilman syvempää pohtimista. Nopea, vaivaton toiminta johtuu siitä, että tilanteessa tarvittavat toiminnot ovat hyvin opittuja eli käyttäjän sisäiset mallit toiminnasta ovat automatisoituneet, tai niin kuin sanonta kuuluu, ne ovat selkäytimessä. Samaan suoritukseen voi kuulua sekä pohdiskelevaa että kokemuksellista toimintaa. (Sinkkonen ym. 2002: 71, Suchman 1987)

Modaliteetin ajallinen ja tilallinen ulottuvuus vaikuttavat siihen, minkälainen toimintatapa on pääosassa kyseisen modaliteetin kautta tapahtuvassa toiminnassa. Voidaan esimerkiksi ajatella, että puhe-käyttöliittymää käyttäessään ihminen toimii enemmän tilannekohtaisessa tilassa, koska dialogin myötä tulevan tiedon lyhyt ajallinen kesto vaatii nopeaa reagoitua. Lisäksi monet puhe-käyttöliittymien yhteydessä käytettävät, keskusteluun liittyvät kognitiiviset prosessit ovat peräisin ihmisten välisestä vuorovaikutuksesta ja ne ovat siten erittäin hyvin opittuja useimmilla ihmisillä. Ihmisen ja koneen erilainen tilannekohtainen reagointi tai ehkä paremminkin koneen kykenemättömyys ihmismäiseen reagointiin saattaa aiheuttaa ristiriitaisia tunteita käyttäjälle.

## **5.6 Graafinen ja puhekäyttöliittymä tukevat toisiaan**

Edellä tarkastellut ominaisuudet olivat monessa suhteessa vastakohtaisia graafisessa ja puhekäyttöliittymässä. Multimodaalisessa käyttöliittymässä toisen modaliteetin vahvoja ominaisuuksia voi käyttää paikkaamaan toisen modaliteetin heikkouksia.

Graafista käyttöliittymää voidaan käyttää tukemaan puhekäyttöliittymää niin, että muistettavaksi tarkoitettu tieto esitetään graafisella näytöllä, jossa se voi säilyä pidempään. LARRI-järjestelmässä korjausohjeet voivat olla näytöllä, jolloin käyttäjä voi tutustua niihin ja niissä mahdollisesti oleviin yksityiskohtaisiin tietoihin rauhassa ja tehdä sen jälkeen toiminnon puheohjatusti.

Samoin puhekäyttöliittymän puutteellista kykyä välittää rinnakkaisista, vertailtavaa tai tilallista informaatiota voidaan tukea graafisella näytöllä. Interact järjestelmässä karttatieto esitetään graafisesti, jolloin käyttäjän voi olla helpompi hahmottaa haluamansa reitti. Vastaavasti LARRI-järjestelmässä näytöllä voidaan esittää rakennekaavioita, joita ei puheen avulla pysty välittämään.

Puhekomentojen muistamisen ongelmaa voidaan helpottaa esittämällä ne näytöllä. CarMMI:n näyttöjen suunnittelussa on esimerkiksi käytetty "speak what you can see" filosofiaa, eli näytöllä olevien valikkojen nimet toimivat myös komentosanoina ja näin käyttäjän on helpompi muistaa tarvittavat sanat. Samoin Interactin karttanäyttö helpottaa oikean paikannimen muistamista.

Myöskin erilaisia palautteita järjestelmän toiminnasta voidaan esittää graafisella näytöllä niin, etteivät ne häiritse puhedialogin kulkua. Esimerkiksi puheentunnistuksen tulokset voidaan esittää näytöllä. Käyttäjä voi korjata tilanteen, jos huomaa näytöllä virheellisen tunnistustuloksen, jolloin erillistä varmistusta ei tarvitse tehdä puhekäyttöliittymän kautta ja sovelluksen käyttö nopeutuu. CarMMI au-

tosovelluksessa puheentunnistuksen tulos näytetään ruudun alareunassa. Samoin ruudulla näytetään puheentunnistuksen päällä olo ja kesto (recognition window) niin, että käyttäjä voi halutessaan katsoa näytöltä, milloin järjestelmälle voi puhua.

Edellisten lisäksi Grasso ym. (1998) esittelevät taulukon 5.1 mukaiset ominaisuudet, jotka täydentävät toisiaan suorakäyttö- ja puhekäyttöliittymissä. Suorakäyttökäyttöliittymä tarkoittaa käytännössä graafisen näytön ja osoitinlaitteen yhdistelmää.

*Taulukko 5.1. Suorakäytön ja puheen toisiaan täydentävät vahvuudet (Grasso ym. 1998).*

<b>Direct Manipulation</b>	<b>Speech Recognition</b>
Direct engagement	Hands/eyes free operation
Simple, intuitive actions	Complex actions possible
Consistent look and feel	Reference does not depend on location
No reference ambiguity	Multiple ways to refer to entities

### **5.7 Puhemodaliteetin ominaisuudet**

Bernsen (2002) ja hänen kollegansa yrittivät luoda sääntöpohjaisen järjestelmän, jonka avulla suunnittelija voisi etsiä automaattisesti sovelluksen vaatimuksiin sopivan modaliteettiyhdistelmän. Järjestelmän perustat osoittautuivat huteriksi ja sen tekeminen sekä käyttäminen niin työlääksi, ettei se ollut käytännössä järkevä. Sääntöpohjaisen lähestymistavan jälkeen he kehittivät metodin, jossa modaliteetin ominaisuuksien perusteella asiantuntija voi tehdä ihmisällyn avulla oivalluksia ja ehdotuksia modaliteetin valinnan suhteen.

Bernsen ja kumppanit testasivat metodologia keräämällä ensin vuodesta 1993 lähtien ilmestyneestä kirjallisuudesta 273 väitettä liittyen puheen ominaisuuksiin ja käyttöön tuotto- ja vastaanottomodaliteettina. Sen jälkeen he analysoivat väitteiden todenperäisyyden sekä et-

sivät modaliteettiteorian pohjalta johdetuista modaliteettiominaisuuksista ne, jotka liittyvät kyseisiin väitteisiin.

Analysointityön pohjalta löytyi 25 modaliteettiominaisuutta, joiden avulla he pystyivät perustelemaan, tukemaan tai korjaamaan lähes kaikki 273 väitettä. Työn pohjalta tehtiin myös SMALTO (the Speech Modality AuxILIary Tool, <http://disc.nis.sdu.dk/smalto/>) työkalu, jonka avulla suunnittelija voi arvioida kannattaako puhetta käyttää halutussa sovelluksessa itsenäisesti tai muiden modaliteettien kanssa (Bernsen 2002).

Taulukossa 5.2 on esitetty Bernsenin löytämät 25 modaliteettiominaisuutta. Osaa ominaisuuksista on käsitelty tässä työssä jo aikaisemmin. Kaikki ominaisuudet (kuten 7 ja 23) eivät suoraan koske puhemodaliteettia, mutta ovat mukana listassa, koska ne liittyvät läheisesti puheen multimodaaliseen käyttöön.

*Taulukko 5.2. Puhemodaliteettiin liittyvät ominaisuudet (Bernsen 2002).*

1	Linguistic input/output modalities have interpretational scope, which makes them eminently suited for conveying abstract information. They are therefore unsuited for conveying high-specificity information including detailed information on spatial manipulation and location. (MP used in the evaluation of 51 claims)
2	Linguistic input/output modalities, being unsuited for specifying detailed information on spatial manipulation, lack an adequate vocabulary for describing the manipulations. (MP used in the evaluation of 10 claims)
3	Arbitrary input/output modalities impose a learning overhead, which increases with the number of arbitrary items to be learned. (MP used in the evaluation of 1 claim)
4	Acoustic input/output modalities are omnidirectional. (MP used in the evaluation of 3 claims)
5	Acoustic input/output modalities do not require limb (including haptic) or visual activity. (MP used in the evaluation of 18 claims)
6	Acoustic output modalities can be used to achieve saliency in low-acoustic environments. They degrade in proportion to competing noise levels. (MP used in the evaluation of 1 claim)
7	Static graphic/haptic input/output modalities allow the simultaneous representation of large amounts of information for free visual/tactile inspection and subsequent interaction. (MP used in the evaluation of 19 claims)
8	Dynamic input/output modalities, being temporal (serial and transient), do not offer the cognitive advantages (wrt. attention and memory) of freedom of perceptual in-

	pection. (MP used in the evaluation of 30 claims)
9	Dynamic acoustic output modalities can be made interactively static (but only small-piece-by-small-piece). (MP used in the evaluation of 12 claims)
10	Speech input/output modalities, being temporal (serial and transient) and non-spatial, should be presented sequentially rather than in parallel. (MP used in the evaluation of 2 claims)
11	Speech input/output modalities in native or known languages have very high saliency. (MP used in the evaluation of 2 claims)
12	Speech output modalities may complement graphic displays for ease of visual inspection. (MP used in the evaluation of 3 claims)
13	Synthetic speech output modalities, being less intelligible than natural speech output, increase cognitive processing load. (MP used in the evaluation of 1 claim)
14	Non-spontaneous speech input modalities (isolated words, connected words) are unnatural and add cognitive processing load. (MP used in the evaluation of 2 claims)
15	Discourse input/output modalities have strong rhetorical potential. (MP used in the evaluation of 7 claims)
16	Discourse input/output modalities are situation-dependent. (MP used in the evaluation of 13 claims)
17	Spontaneous spoken labels/keywords and discourse input/output modalities are natural for humans in the sense that they are learnt from early on (by most people and in a particular tongue and, possibly, accent). (Note that spontaneous keywords and discourse must be distinguished from designer-designed keywords and discourse, which are not necessarily natural to the actual users.) (MP used in the evaluation of 29 claims)
18	Notational input/output modalities impose a learning overhead, which increases with the number of items to be learned. (MP used in the evaluation of 4 claims)
19	Analogue graphics input/output modalities lack interpretational scope, which makes them eminently suited for conveying high-specificity information. They are therefore unsuited for conveying abstract information. (MP used in the evaluation of 5 claims)
20	Direct manipulation selection input into graphic output space can be lengthy if the user is dealing with deep hierarchies, extended series of links, or the setting of a large number of parameters. (MP used in the evaluation of 10 claims)
21	Haptic deictic input gesture is eminently suited for spatial manipulation and indication of spatial location. It is not suited for conveying abstract information. (MP used in the evaluation of 25 claims)
22	Linguistic text and discourse input/output modalities have very high expressiveness. (MP used in the evaluation of 2 claims)
23	Images have specificity and are eminently suited for representing high-specificity information on spatio-temporal objects and situations. They are therefore unsuited for conveying abstract information. (MP used in the evaluation of 4 claims)
24	Text input/output modalities are basically situation-independent. (MP used in the evaluation of 10 claims)
25	Speech input/output modalities, being physically realised in the acoustic medium, possess a broad range of acoustic information channels for the natural expression of information. (MP used in the evaluation of 8 claims)

Edellisten lisäksi Shneiderman (2000) esittää mielenkiintoisen väitteen, jonka mukaan puhe varaa lyhytkestoista- ja työmuistia itselleen ja kuluttaa näin ongelmanratkaisuun ja muistiin palauttamiseen tarvittavaa kognitiivista kapasiteettia. Haptis-visuaalinen suorakäyttö ei varaa Shneidermanin mukaan muistikapasiteettia samalla tavalla. Hän on myös esittänyt joitain kokeellisia tuloksia multimodaalisista vuorovaikutustilanteista, jotka antavat suuntaa siitä, että puheen käyttö sovelluksen ohjaamiseen heikentäisi muistamista hiiren käyttöön verrattuna.

### **5.8 Puhekäyttöliittymien tekniset erityispiirteet**

Puheen tunnistus ja tuottaminen on teknisesti vaativaa, eikä tekniikkaa voida käytännössä jättää huomioimatta käyttöliittymää suunniteltaessa. Tässä kappaleessa esitellään muutamia keskeisimpiä teknisiä rajoituksia.

Suomalaisille samoin kuin muille pienten kielten edustajille eräs rajoitus on se, että varsinkin kaupallinen puhekäyttöliittymätekniikka on suunniteltu tukemaan etupäässä englantia ja muutamaa muuta suurempaa kieltä. Pienille kielille valmiita tuotteita on huomattavasti vähemmän saatavana.

#### **5.8.1 Puheentunnistuksen rajoitukset**

Puhe voidaan luokitella neljään eri luokkaan: jatkuva puhe, puhe, jossa sanojen välissä on lyhyet tauot, erilliset sanat ja kirjain kerrallaan lausuminen. Lisäksi voidaan erotella erilaisia puhetyylejä, kuten sisäluku (vrt. radion uutistenlukija), spontaani puhe (vrt. tavallinen puhelinkeskustelu), ylikorostettu puhe. Puheen tyyppi ja tyyli vaikuttavat olennaisesti puheen tunnistuksen onnistumiseen ja tarkkuuteen. (Gibbon ym. 2000)



Puheentunnistimet voidaan jakaa kahteen luokkaan käytetyn puhe-  
tyylin perusteella. Erillisen puheen tunnistimet, joiden kanssa puhu-  
jan täytyy pitää taukoa sanojen tai kirjainten välillä, jotta ne erot-  
tuisivat selkeästi toisistaan ja jatkuvan puheen tunnistimet, joille  
käyttäjät voivat puhua luonnollisemmin ilman liioiteltuja taukoja sa-  
nojen välissä. Jatkuvan puheen tunnistus on teknisesti vaikeampaa  
ja siten häiriöalttiimpaa, mutta vastaavasti se mahdollistaa monessa  
sovelluksessa jouhevamman käyttöliittymän. (Gibbon ym. 2000)

Puheentunnistusjärjestelmät voidaan jakaa kahteen luokkaan myös  
sen perusteella täytyykö niitä opettaa. Toiset järjestelmät täytyy  
aluksi opettaa tunnistamaan kunkin käyttäjän oma puhe. Toisiin jär-  
jestelmiin tilastollinen malli on rakennettu valmiiksi esimerkiksi  
käyttäen useilta käyttäjiltä kerättyä puhemateriaalia, jolloin järjes-  
telmät tunnistavat useamman käyttäjän puhetta ilman, että järjes-  
telmää täytyisi opettaa kunkin käyttäjän puheelle erikseen. Ensien  
mainitut järjestelmät ovat usein tarkempia tunnistustuloksissaan,  
toisaalta jälkimmäiset ovat huomattavasti helppokäyttöisempiä mo-  
nissa käyttötilanteissa.

Suurin osa nykyisistä puheentunnistinjärjestelmistä toimii siten, että  
etukäteen määritellään sanasto, jota järjestelmä kuuntelee ja antaa  
sitten vastauksena todennäköisimmät sanat kuunneltavasta sanas-  
tosta. Sanasto voi muodostua yksittäisistä sanoista tai lauseen ta-  
paisista peräkkäisten sanojen muodostamista ketjuista. Yleisesti ot-  
taen tunnistuksen tarkkuus heikkenee sitä mukaan, kun sanasto  
kasvaa.

Toinen sanastoon liittyvä rajoitus tulee siitä, että järjestelmän kan-  
nalta lähekkäin olevat sanat saattavat sekoittua keskenään. Keski-  
määrin pitkät sanat erottuvat toisistaan paremmin kuin lyhyet. Aina  
ei ole etukäteen selvää kuinka järjestelmä "kuulee" sanat. Jotkut  
sanat saattavat olla käytetyn tilastollisen mallin kannalta lähellä toi-

siaan, vaikka ne ihmisen mielestä kuulostavat aivan erilaiselta. Tämä täytyy ottaa huomioon sovellusta kehitettäessä niin, että sanasto voidaan muokata kehityksen aikana ja sen jälkeen, jos ongelmia esiintyy.

Ihmisten välisessä keskustelussa on olemassa tiettyjä kohtia, jolloin keskustelun vuoro vaihtuu luontevimmin, mutta silti usein joku keskustelun osapuoli keskeyttää toisen sanoakseen oman asiansa. Puhelukäyttöliittymissä tulee vastaan samanlaisia tilanteita, joissa käyttäjä haluaisi alkaa puhumaan, vaikka järjestelmä vielä antaa palautetta. Toiset järjestelmät tukevat "barge-in":ksi nimettyä ominaisuutta, jonka avulla järjestelmä voi aloittaa kuuntelun, kun käyttäjä sanoo jotain kesken palautteen. Barge-in on käytännössä teknisesti vaativaa toteuttaa ja sen toimivuus saattaa riippua esimerkiksi taustamelun määrästä. Sen takia on suunnittelun kannalta tärkeää miettiä, missä tapauksissa toimintoa käytetään ja milloin se on kytketty pois päältä.

Taustahäly saattaa muutenkin heikentää puheentunnistusjärjestelmän toimintaa. Nykyaikaisilla signaalinkäsittely- ja suodatusmenetelmillä taustahälyn vaikutusta voidaan pienentää huomattavasti.

### **5.8.2 Puheentuoton rajoitukset**

Puhesyntetisaattorit toimivat siten, että niille syötetään puhuttavaksi haluttavat sanat yleensä tekstimuodossa ja puhesyntetisaattori muuttaa syötteen ääneksi. Ihmisten välisessä puheessa on muutaakin informaatiota kuin pelkät sanat. Esimerkiksi puheen rytmillä, sanojen ja tavujen painotuksella, nopeuden ja äänenkorkeuden muutoksilla eli prosodialla on olennainen tehtävä viestintätilanteessa. Puhesyntetisaattorille voidaan antaa tekstin lisäksi ohjausinformaatiota, jonka avulla edellä mainittuja ilmiöitä voidaan jäljitellä tai syntetisaattori voi yrittää päätellä oikeat painotukset tekstin rakenteen perusteella.

Vaikka nykyiset puhesyntetisaattorit ovat jo melko hyviä, eivät ne kuulosta täysin luonnollisilta, mikä aiheuttaa sen, että käyttäjän on vaikeampi ymmärtää syntetisoitua kuin luonnollista puhetta. Tämä korostuu erityisesti häiriöisessä kommunikaatitilanteessa. Joissain tapauksissa käyttäjät suosivat syntetisoitua ääntä, koska se erottuu muista (Cohen & Oviatt 1994. Weinschenkin & Barkerin 2000, mukaan).

Puhesynteesin rajoituksia on pyritty kiertämään käyttämällä äänitteitä. Valmiiksi nauhoitetut lauseet tai sanat soitetaan käyttäjälle sellaisenaan. Tällä tavoin saadaan puheen laatu hyväksi ja ymmärrettäväksi. Haittapuoli synteesiin verrattuna on menetelmän jäykkyys. Valmiiksi nauhoitettujen puheenpätkien avulla on hankala toteuttaa muuttuvia palautteita kuten puhelinnumeroita ilman, että puheeseen tulee omituisen kuuloisia epäsujuvuuskohtia.

### **5.8.3 Dialogin hallinnan malleja**

Olenainen osa puhekäyttöliittymän toteutusta on vuoropuhelun eli dialogin hallinta, jonka avulla käyttäjän ja tietokoneen puheenvuoroja koordinoidaan. Käytettävissä oleva tekninen dialoginhallintamalli vaikuttaa huomattavasti käyttöliittymän suunnittelumahdollisuuksiin.

Yksinkertaisimmillaan vuoropuhelu sovelluksen kanssa menee aina saman kiinteän kaavan mukaan, jolloin tekninen toteutus on helppompaa, mutta käyttäjä saattaa kokea sovelluksen joustamattomana.

Käyttöä voidaan tehdä sujuvammaksi käyttämällä teknisesti hieman monimutkaisempia oikopolku- tai lomakemalleja. Oikopolkujen avulla käyttäjä voi siirtyä tiettyä komentosanaa käyttäen suoraan johonkin toiseen osaan puhesovellusta kulkematta ensin koko valikko-

hierarkiaa läpi. Lomakemallissa puolestaan ajatellaan, että käyttäjä ikään kuin täyttää lomaketta puheesityönteillään. Käyttäjä voi käytännössä täyttää useita kohtia lomakkeessa samalla lausahduksella ja järjestelmä voi sen jälkeen pyytää käyttäjää antamaan mahdollisesti puuttuvat tiedot. Myös järjestelmän antamaa palautetta voidaan muunnella niin, ettei käyttäjä kyllästy aina samanlaisena toistuvaan vastaukseen.

Tämän hetken parhaissa dialoginhallintajärjestelmissä pyritään mallintamaan ihmisten välistä vuorovaikutusta. Niissä otetaan huomioon esimerkiksi keskustelun aikana kertyneet ja aihealueeseen liittyvät tiedot, joiden avulla vuoropuhelusta voidaan tehdä sulavampaa (katso kappale 3.7.3 - Interact).

Multimodaalisissa järjestelmissä dialoginhallinta kohtaa uusia haasteita, kun syötteitä ja palautteita täytyy koordinoita eri modaliteettien välillä.

## 6. Modaliteettien valinta ja yhdistäminen

Multimodaalisessa järjestelmässä on vaihtoehtoisia kanavia välittää tietoa käyttäjän ja koneen välillä. Suunnittelijan tehtävänä on valita modaliteetin tai modaliteettien yhdistelmät niin, että ne tukevat parhaiten käyttäjää tehtävän suorittamisessa. Suunnittelussa täytyy ottaa huomioon tehtävissä tarvittavan vuorovaikutuksen ja informaation esitystapojen asettamat vaatimukset, käyttäjän ominaisuudet ja mieltymykset sekä käyttötilanne.

Modaliteettien valinta ei ole yksinkertainen tehtävä. Modaliteetteja on useita ja niiden yhdistelmiä valtava määrä. Valmista teoreettista viitekehystä, jonka avulla valinnan voisi tehdä ei vielä ole. Käytännön tutkimuksen kautta on kertynyt tietoa toimivista yhdistelmistä tiettyihin käyttötarkoituksiin, mutta tietoa on niin vähän ettei, kattavaa ohjetta ole vielä olemassa. Raisamo (1999) luettelee tutkimusalueita, jotka pitäisi kartoittaa, jotta ohjeen laatiminen olisi mahdollista:

- Kuinka aivot toimivat ja mitä modaliteetteja voitaisiin parhaiten käyttää multimodaalisen vuorovaikutuksen mahdollistamien synergiaetujen saavuttamiseksi.
- Milloin multimodaalinen järjestelmä on parempi kuin yksimodaalinen.
- Mitkä modaliteetit sopivat parhaiten tiettyyn vuorovaikutustehtävään.
- Miten vuorovaikutuslaitteet tulisi valita suhteessa modaliteetteihin ja miten niitä tulisi käyttää ja kehittää.

Tämän työn kannalta modaliteetit on rajattu käytännössä työn rajauksen myötä. Katso kohdat 3.4 ja 3.5.

Vaikka modaliteetit on rajattu, jää vielä monta avointa kysymystä. Miten valitut modaliteetit voi yhdistää ja miten valita sopiva modali-

teettiyhdistelmä kuhunkin tilanteeseen? Kuten edellä todettiin, kattavia ohjeita tai teoriaa ei ole vielä olemassa, joten asiaa täytyy pohtia loogisen analyysin ja käytännön esimerkkien avulla. Ilmiön monimutkaisuudesta johtuen, esimerkkisovelluksista saadun tiedon yleistämiseen täytyy suhtautua kriittisesti (Bernsen 2002).

Schomaker ym. (1995:49) esittävät modaliteettien yhdistämisen vaikutuksista kaksi hypoteesia, jotka kuvaavat hyvin yhdistämisen ongelmaa:

"Hypothesis (1): The combination of different human output channels is functionally interesting because it effectively increases the bandwidth of the human machine channel. Examples are the combination of bimanual teleoperation with added speech control. A well-known example is the typing on a keyboard where key sequences which require alternating left and right hand keying are produced 25% faster (50 ms) [250] than a within-hand keying sequence. The reason is thought to reside in the independent, parallel control of left and right hand by the right and left hemispheres, respectively. It may be hypothesized that other forms of multimodality profit in a similar way, if their neural control is sufficiently independent between the modalities involved.

Hypothesis (2): The alternative hypothesis goes like this: Adding an extra output modality requires more neurocomputational resources and will lead to deteriorated output quality, resulting in a reduced effective bandwidth. Two types of effects are usually observed: (a) a slowing down of all the output processes, and (b) interference errors due to the fact that selective attention cannot be divided between the increased number of output channels. Examples are writing errors due to phonemic interference when speaking at the same time, or the difficulty people may have in combining a complex motor task with speaking such as in simultaneously driving a car and speaking, or playing a musical instrument and speaking at the same time. This type of reasoning is typical for the cognition-oriented

models of motor control and may provide useful guidelines for experimentation. ”

Hypoteesit koskevat tuottomodaliteetteja, mutta vastaavat voidaan varmasti esittää myös vastaanottomodaliteettien osalta. Edellä esitetyt hypoteesit koskevat modaliteettien samanaikaista käyttöä. Toisissa sovelluksissa modaliteetteja ei välttämättä käytetä yhtäaikaan vaan esimerkiksi peräkkäin.

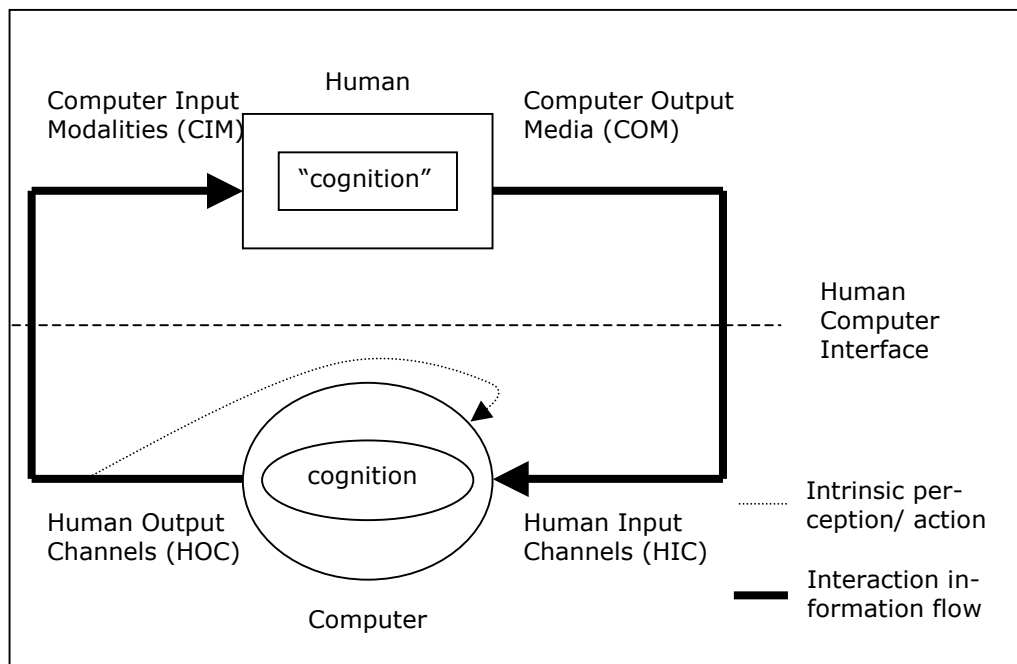
### **6.1 Vuorovaikutustilanteen malli**

Vuorovaikutuksen analysointiin tarvitaan jonkinlainen malli. HCI-piireissä yleinen tapa on määrittää rajapinnan paikka ihmisen ja koneen välille. Tästä seuraa, että rajapinnan kautta liikkuvasta informaatiosta (eli syötteistä ja palautteista) tulee keskeinen osa vuorovaikutusta.

Multimodaalisessa järjestelmässä syötteet ja palautteet jakautuvat usean modaliteetin kesken. Seuraavissa kappaleissa pohditaan, miten syötteet ja palautteet voivat jakautua eri modaliteettien välillä ja toisaalta ilmiöitä, jotka vaikuttavat siihen, miten jako olisi järkevää tehdä.

Kirjallisuudessa on esitetty kuvan 6.1 mukainen malli multimodaalisesta vuorovaikutuksesta (Schomaker ym. 1995, Gibbon ym. 2000).

Mallissa on vähintään kaksi fyysisesti erillistä toimijaa, ihminen ja tietokone, jotka voivat välittää tietoja toisilleen useita kanavia käyttäen rajapinnan läpi.



Kuva 6.1. Multimodaalisen vuorovaikutustilanteen malli (Schomaker ym. 1995)

Mallissa on kaksi käyttäjän näkökulmasta kuvattua prosessia, havaintoprosessi (Perception process) ja ohjausprosessi (Control process). Havaintoprosessi viittaa ihmisen tietokoneen palautteesta tekemiin havaintoihin ja siihen kuuluu tietokoneen tuottomodaliteetit (Computer Output Media - COM) ja ihmisen vastaanottomodaliteetit (Human Input Channels - HIC). Ohjausprosessi viittaa tietokoneen tekemiin tulkintoihin ihmisen ohjaustoimista ja siihen kuuluu ihmisen tuottomodaliteetit (Human Output Channels, HOC) ja tietokoneen vastaanottomodaliteetit (Computer Input Modalities, CIM).

Malliin kuuluu kaksi vuorovaikutuskehää. Sisäinen palautekehä (Internal Perception/Action loop), joka viittaa ihmisen omaan kykyyn tarkkailla omaa toimintaansa ja ulkoinen palautekehä (Interaction Information Flow), joka kuvaa tiedon kulkua vuorovaikutustilanteessa.

Lisäksi mallissa on älykkyyteen viittaava komponentti kummankin toimijan sisällä. Ihmisellä se koostuu luonnollisesti aivoista ja niissä



toimivista kognitiivisista prosesseista (cognition) ja tietokoneessa laskennallisista algoritmeista ("cognition").

Mallissa ei suoraa erotella eri modaliteetteja toisistaan, vaan ne kuuluvat samaan nippuun.

Malli on yksinkertaisuudessaan havainnollinen, mutta toisaalta siitä puuttuu monia vuorovaikutukseen olennaisesti vaikuttavia tekijöitä. Mallista saa sen kuvan, että kyseessä on suljettu järjestelmä, mikä ei todellisessa käyttötilanteessa pidä paikkaansa. Käyttäjä tuo aina yksilöllisen kokemuksensa mukaan vuorovaikutukseen aikaisemman kokemuksensa kautta.

Ympäristön vaikutusta ei ole myöskään huomioitu mitenkään, vaikka se on erittäin olennainen monissa multimodaalisissa sovelluksissa. Ympäristö vaikuttaa aina ihmiseen ja voidaan ajatella, että siksi sen pitäisi vaikuttaa myös tietokoneeseen

## **6.2 Yhdistämisen tavat**

Aluksi voidaan miettiä yleisellä tasolla modaliteettien mahdollisia eri yhdistämistapoja. Martin ym. (1995) ovat esittäneet luokittelun, joka erottelee kuusi eri tapaa, joilla modaliteetit voivat toimia yhteistyössä (Gibbon ym. 2000: 140).

### **6.2.1 Toisiaan täydentäen (Complementarity)**

Samaan tehtävään liittyvät osasyötteet tai -palautteet välitetään käyttäen eri modaliteetteja. Näin voidaan käyttää aina sopivinta modaliteettia tietyn tyyppisen tiedon välittämiseen ja modaliteetit täydentävät toisiaan. Esimerkiksi joukkoliikenneneuvontapalvelussa kysymys "Millä bussilla pääsen täältä tänne?" esitetään puheena ja maantieteelliset paikat, joihin kysymys viittaa, osoitetaan kosketus-herkältä karttanäytöltä sormella painamalla. Kuten kappaleessa 5.3

todettiin, tilalliset viittaukset on usein helpompi tehdä käyttäen graafista- kuin puhekäyttöliittymää.

### **6.2.2 Toisiaan vahvistaen (Redundancy)**

Sama tieto voidaan välittää useamman modaliteetin kautta yhtäaikaan, jolloin eri modaliteettien kautta tulleet syötteet tai palautteet vahvistavat toisiaan. Tästä on hyötyä esimerkiksi häiriöisessä ympäristössä, jolloin viestin merkitys voidaan tulkita varmemmin. Samoin sellaisissa tehtävissä, joissa sallitaan hyvin vähän virhetulkintoja. Esimerkiksi puhuvaa päätä voidaan käyttää tuottamaan puhetta vastaava visuaalinen palaute huulenliikkeiden avulla niin, että puheen tunnistaminen helpottuu.

### **6.2.3 Tasa-arvoisesti (Equivalence)**

Sama tieto voidaan välittää käyttäen useampaa eri modaliteettia. Käyttäjä (tai järjestelmä) voi valita mieltymyksensä tai käyttötilanteen mukaan sen, mitä modaliteettia kulloinkin käyttää. Esimerkiksi joukkoliikenneneuvontapalvelussa käyttäjä voi valita, esittääkö hän matkan päätepisteet puheena (Käyttäjä kysyy: "Millä bussilla pääsen Arabiasta Kallioon?") vai kosketusnäyttöä käyttäen (Käyttäjä kysyy: "Millä bussilla pääsen täältä tänne?" ja osoittaa sormella Arabian ja Kallion kartalta).

### **6.2.4 Erikoistuneesti (Specialisation)**

Tietty tieto järjestelmässä välitetään aina käyttäen samaa modaliteettia. Vastakohta tasa-arvoiselle yhdistämisen tavalle. Suunnittelija voi valita mielestään parhaan mahdollisen modaliteetin kuhunkin tehtävään. Erikoistumista voidaan ajatella myös käyttäjän näkökulmasta, jos käyttäjä suosii tehtävän suorittamiseen jotain tiettyä modaliteettia muiden sijaan käyttötilanteessa, jossa on tarjolla enemmän kuin yksi tasa-arvoinen modaliteetti. Esimerkiksi informaatiokioskissa käytettävät palvelut valitaan aina graafista käyttö-

liittymää ja kosketusnäyttöä käyttäen, vaikka varsinaisissa palveluissa on mahdollista käyttää myös puhetta ohjaukseen.

### **6.2.5 Yhtäaikaisesti (Concurrency)**

Eri modaliteetteja käytetään samanaikaisesti ohjaamaan eri tehtäviä järjestelmässä. Arkielämän esimerkki vastaavasta tilanteesta olisi dokumentin muokkaus tekstinkäsittelyohjelmalla ja samaan aikaan (kaiutin)puhelimeen puhuminen. Tällaisen järjestelmän käyttö saattaa olla kognitiivisesti rasittavaa, koska käyttäjä joutuu jakamaan tarkkaavaisuutensa moneen suuntaan yhtäaikaa.

### **6.2.6 Modaliteettia muuttaen (Transfer)**

Toisen modaliteetin kautta välitetty tieto käynnistää toiminnon toisessa modaliteetissa. Käytännössä siis tehtävään liittyvät syötteet ja palautteet kulkevat eri modaliteettien kautta. Esimerkiksi puhekomennolla voi vaihtaa näytön sisältöä autotietokonesovelluksessa.

Kaikki edellä olevat yhdistämisen tavat ovat mahdollisia käytännön sovelluksissa. Riippuu sovelluksesta, mikä on käytännössä paras tapa.

### **6.2.7 Peräkkäin vs. rinnakkain**

Nigay ja Coutaz (1993) esittelevät kuvan 6.2 mukaisen vaihtoehtoisen luokittelun modaliteettien yhdistämiselle. Luokittelu on tehty järjestelmän näkökulmasta, mutta samaa voidaan soveltaa myös käyttäjän näkökulmasta.

		USE OF MODALITIES	
		Sequential	Parallel
FUSION	Combined	ALTERNATE	SYNERGISTIC
	Independent	EXCLUSIVE	CONCURRENT
		Meaning No Meaning	Meaning No Meaning
LEVELS OF ABSTRACTION			

Kuva 6.2. Multimodaalisten järjestelmien suunnitteluavaruus (Nigay & Coutaz 1993)

Tässä luokittelussa otetaan ajallinen ulottuvuus mukaan, Martinin ja kumppaneiden luokittelussa se oli mukana implisiittisesti. Jos katsotaan edellisessä kappaleessa kuvattuja yhdistämisen tapoja voidaan huomata, että osassa tavoista (toisiaan täydentäen, toisiaan vahvistuen, yhtäaikaaisesti) eri modaliteetteja käytetään rinnakkain ja osassa (tasa-arvoisesti, erikoistuneesti, modaliteettia muuttaen) peräkkäin tai vuorotellen.

Modaliteettien peräkkäisessä käytössä täytyy tehtävien alku- ja loppupalautteiden olla sellaiset, että käyttäjä tietää, mitä modaliteettia pitäisi käyttää seuraavaksi ja milloin modaliteetin vaihto voi tapahtua.

### 6.3 Modaliteettien tahdistus

Multimodaalisissa käyttöliittymissä modaliteetteja käytetään usein yhtä aikaa, jolloin nousee esiin kysymys modaliteettien tahdistuksesta eli synkronoinnista keskenään. Jotta käyttäjä tai järjestelmä voisi havaita eri modaliteettien kautta välittyvien asioiden kuuluvan yhteen, täytyy niiden tapahtua ajallisesti suhteellisen lähekkäin.

Tahdistuksen tarkkuuden vaatimus on kuitenkin käytännössä vahvasti riippuvainen esitettävästä tai syötettävästä informaatiosta. Esimerkiksi videokuvassa, jossa ihminen puhuu ja hänen äänensä kuuluu, ihminen havaitsee eron, jos ääni tulee yli 130 ms ennen kuvaa tai jos kuva tulee yli 260 ms ennen ääntä eli ihminen on tottunut havaitsemaan kuvan ennen ääntä. Jos kyseessä on terävät äänneet, aikaikkuna pienenee entisestään (130 ms -> 75 ms, 260 ms -> 188 ms). (Dixon & Spitz 1980. Schomaker ym. 1995 mukaan.) Myös Hollier ym. (1999) huomasivat, että sisällöllä on merkitystä synkronointivirheiden havaitsemisessa. Ihminen on erityisen tarkka huomaamaan puhuvien kasvojen ja puheäänien eriaikaisuuden. Jos visuaaliset vihjeet eivät ole tarkkoja tai sellaisia, että ihmiset eivät ole tottuneet havaitsemaan niitä, esim. lintu laulaa kaukana puussa tai piirrettyjen elokuvien hahmojen suunliikkeet, ei kuvan ja äänen synkronoinninkaan tarvitse olla niin tarkka.

Edellä kuvatut tulokset koskivat audiovisuaalisen informaation vastaanottotilannetta. Tahdistukselle ei voida välttämättä esittää yhtä yksityiskohtaisia yleistettävissä olevia raja-arvoja vuorovaikutteisiin tilanteisiin, joissa käyttäjä joutuu mahdollisesti ohjaamaan tarkkaavaisuuttaan eri kohteisiin tehtävän aikana. Käytännössä raja-arvot joudutaan määrittämään tapauskohtaisesti.

Käyttäjien pitäisi pystyä hyödyntämään modaliteetteja omien mieltymystensä mukaan. Käytännössä käyttäjät eivät välttämättä käytä modaliteetteja ajallisesti täysin yhtäaikaan rinnakkaisiksikaan luokitelluissa käyttötapauksissa. Oviattin (1999) testatessa kynä- ja puhekäyttöliittymät yhdistävää multimodaalista sovellusta, hän huomasi, että kynämerkintä edeltää yleensä puhekomentoa suuressa osassa tapauksista niin, etteivät ne tapahdu lainkaan päällekkäisesti, vaikka käyttäjän näkökulmasta modaliteetit ovat rinnakkain käytössä. Käyttäjien välillä esiintyi myös yksilökohtaisia eroja siinä,

kuinka lähekkäinen tai päällekkäin he käyttivät modaliteetteja. Vaihdelut täytyy ottaa huomioon järjestelmää suunniteltaessa.

#### **6.4 Käyttäjän mukaan**

Kuten luvussa 4.1 todettiin käyttäjät ovat monella tavalla erilaisia ja multimodaalisuuden avulla voidaan tukea käyttäjien vaihtelevia tapoja tai mahdollisuuksia käyttää sovelluksia.

Seuraavassa on arvioitu SMALTO:sta löydettyjen väitteiden, kappaleessa 5.7 esiteltyjen 25:n puhemodaliteetin ominaisuuden ja intuition perusteella graafisen ja puhekäyttöliittymän soveltuvuutta erilaisille käyttäjille.

Tuottomodaliteettien kohdalla täytyy ottaa huomioon, että käyttäjän täytyy yleensä pystyä tulkitsemaan saman modaliteetin palautetta voidakseen syöttää tietoa. Esimerkiksi kirjoitettu teksti olisi hyvä pystyä näkemään, jotta tietää tehneensä virheetöntä kirjoitusta, hiiren kohdistin täytyy nähdä, jotta sen voisi kohdistaa ja puhekäyttöliittymän palautteet olisi hyvä pystyä kuulemaan, jotta dialogin kulua pystyisi seuraamaan. Multimodaalisuus mahdollistaa syötteiden ja palautteiden hajottamisen eri modaliteetteihin.

##### **6.4.1 Tuottomodaliteetit**

Graafisen käyttöliittymän arvioinnissa on mukana siihen olennaisesti liittyvä osoittamiseen tai tekstinsyöttöön tarvittava haptinen modaliteetti.

<b>Graafinen sopii</b>
Käyttäjille, jotka eivät pysty puhumaan
Käyttäjille, jotka eivät ymmärrä hyvin sovelluksessa käytettyä kieltä. Graafinen käyttöliittymä tarjoaa enemmän aikaa ymmärtämiseen.
Käyttäjille, joilla ei ole käsien motorisia vaikeuksia.
Visuaalispainotteisille ihmisille
Kinesteettispainotteisille ihmisille

<b>Graafinen ei sovi</b>
Sokeille käyttäjille
Heikkonäköisille käyttäjille
Lukihäiriöisille käyttäjille, sovelluksesta riippuen
Käyttäjille, jotka eivät osaa lukea ja/tai kirjoittaa. (Siinä tapauksessa, että sovelluksessa käytetään tekstiä)
Käyttäjille, joilla ei ole käsiä
Käyttäjille, joilla on käsien motorisia ongelmia
Auditiivispainotteisille ihmisille

<b>Puhe sopii</b>
Sokeille käyttäjille
Heikkonäköisille käyttäjille
Käyttäjille, jotka eivät osaa lukea tai kirjoittaa (esim. lapset)
Käyttäjille, jotka eivät pysty käyttämään käsiään lainkaan tai riittävän tarkasti
Käyttäjille, jotka pitävät enemmän puhumisesta kuin muista modaaliteeteista
Lukihäiriöisille käyttäjille
Auditiivispainotteisille käyttäjille

<b>Puhe ei sovi</b>
Käyttäjille, jotka eivät pysty puhumaan
Käyttäjille, jotka pystyvät puhumaan huonosti
Käyttäjille, jotka eivät osaa sovelluksessa käytettyä kieltä
Käyttäjille, joilla on vahva aksentti. (Aksentti voi aiheuttaa ongelmia puheentun- nituksen kanssa)
Käyttäjille, jotka eivät pidä koneelle puhumisesta
Kuuroille käyttäjille
Heikkokuuloisille käyttäjille
Visuaalispainotteisille ihmisille
Kinesteettispainotteisille ihmisille

#### 6.4.2 Vastaanottomodaliteetit

<b>Graafinen sopii</b>
Kuuroille käyttäjille
Huonokuuloisille käyttäjille
Käyttäjille, jotka eivät ymmärrä hyvin sovelluksessa käytettyä kieltä. Graafinen käyttöliittymä tarjoaa enemmän aikaa ymmärtämiseen.
Visuaalispainotteisille ihmisille

<b>Graafinen ei sovi</b>
Sokeille käyttäjille
Heikkonäköisille käyttäjille
Henkilöille, jotka eivät osaa lukea (esim. lapset)
Lukihäiriöisille käyttäjille, sovelluksesta riippuen
Auditiivispainotteisille ihmisille

<b>Puhe sopii</b>
Sokeille käyttäjille
Heikkonäköisille käyttäjille
Henkilöille, jotka eivät osaa lukea (esim. lapset)
Lukihäiriöisille käyttäjille
Auditiivispainotteisille ihmisille



<b>Puhe ei sovi</b>
Kuuroille käyttäjille
Huonokuuloisille käyttäjille.
Henkilöille, jotka ymmärtävät huonosti käytettyä kieltä
Visuaalispainotteisille ihmisille

On hyvä huomata, että edellä luetellut rajoitukset ovat eri tasoisia. Esimerkiksi auditiivispainotteinen oppimistyyli tai lukihäiriö eivät esitä samalla tavalla graafisen sovelluksen käyttöä kuin näköaistin puuttuminen.

### 6.4.3 Erityisryhmät

Edellä olevia taulukoita tarkasteltaessa voidaan huomata, että monet modaliteettien rajoituksista koskevat erilaisia erityisryhmiä, kuten kuulo- ja näkövammaisia sekä motorisesti rajoittuneita. Erityisryhmien edustajat ovat usein erittäin motivoituneita käyttämään erikoisiakin käyttöliittymäratkaisuja pystyäkseen hyödyntämään niitä nykytekniikan tarjoamia palveluita, joita tavallisille käyttäjille on tarjolla.

Erilaisten yksimodaalisten vuorovaikutustapojen lisääminen parantaa jo sinänsä erityisryhmien mahdollisuuksia käyttää palveluita. Multimodaalisuus mahdollistaa lisäksi syötteiden ja palautteiden hajottamisen eri modaliteetteihin. Näin on mahdollista ottaa erityisryhmien tarpeet vielä monipuolisemmin huomioon.

Esimerkiksi käyttöliittymä, jossa (piste)näppäimistöllä kirjoitettu teksti luetaan takaisin puhesyntetisaattorilla, voi olla suureksi avuksi näkövammaiselle. Markkinoilla oleva Memona Plus muistiinpanolaite toimii juuri edellä kuvatulla tavalla. Vastaavasti niille kuulonsa menettäneille, jotka pystyvät puhumaan, voitaisiin ajatella käyttöliittymää, jossa puheen avulla annettujen syötteiden palautteet tulevat graafiselle näytölle.

#### **6.4.4 Design for all**

”Design for all” tai ”Universal design” ajattelun mukaisesti erityisryhmien lähtökohdista suunnitellut sovellukset saattavat olla helpokäyttöisempiä myös valtaväestölle (Keinonen 2000). Multimodaalisuus avaa uusia mahdollisuuksia ”Design for all” -tyyliseen käyttöliittymäsuunnitteluun.

Puhekäyttöliittymän ja graafisen käyttöliittymän yhdistelmän suuri etu on se, että modaliteetit tukevat toisiaan paikkaamalla toistensa heikkouksia monella tavalla. Erityiskäyttäjien näkökulmasta asiaa tarkasteltaessa tulee kuitenkin helposti eteen tilanne, että käyttö on mahdotonta, jos kumpaakin modaliteettia täytyy käyttää yhtä aikaa. Näkövammaisen ei voi käyttää sovellusta, jossa joutuu katsomaan, eikä kuulovammaisen vastaavasti sellaista, jossa tarvitaan kuuloaitia.

Jos halutaan, että mahdollisimman moni käyttäjä pystyisi käyttämään sovellusta, pitäisi sovelluksen kaikkien toimintojen olla käytettävissä mahdollisimman monen modaliteetin kautta. Toisin sanoen, kappaleessa 6.2 esiteltyjä määritelmiä käyttäen, tasa-arvoisesti tai jossain määrin toisiaan vahvistaen yhdistetyt modaliteetit toimivat tässä tapauksessa paremmin kuin toisiaan täydentävästi tai erikoistuneesti yhdistetyt.

Tavallisille käyttäjille modaliteettien yhdistäminen tasa-arvoisesti tai toisiaan vahvistaen antaa mahdollisuuden modaliteettien valintaan omien mieltymystensä tai vaihtelevien käyttötilanteiden mukaan, jolloin ”Design for all” -ideologia voi toteutua mainiosti.

#### **6.5 Käyttötilanteen mukaan**

Luvussa 4.2 kuvailtiin tietokonesovellusten yhä moninaisemmiksi muuttuvia käyttötilanteita. Multimodaalisutta voidaan hyödyntää sovellusten sovittamiseksi paremmin vaihteleviin käyttötilanteisiin.

Seuraavassa on arvioitu SMALTO:sta löydettyjen väitteiden, kappa-  
leessa 5.7 esiteltyjen 25:n puhemodaliteetin ominaisuuden ja intui-  
tion perusteella graafisen ja puhekäyttöliittymän soveltuvuutta eri-  
laisiin käyttötilanteisiin.

### 6.5.1 Tuottomodaliteetit

Graafisen käyttöliittymän arvioinnissa on mukana siihen olennaisesti  
liittyvä osoittamiseen tai tekstinsyöttöön tarvittava haptinen modali-  
teetti.

<b>Graafinen sopii</b>
Liikkuviin käyttötilanteisiin, koosta ja syöttölaitteesta riippuen
Meluisiin ja kohinaisiin käyttötilanteisiin
Käyttötilanteisiin, joissa muut voivat häiriintyä äänestä
Käyttötilanteisiin, joissa käyttäjä haluaa säilyttää yksityisyytensä
Käyttötilanteisiin, joissa jotain sovelluksen tietoa pitää pystyä tarkastelemaan ajasta riippumatta
Käyttötilanteisiin, joissa viestistä halutaan peittää luonnolliset reaktiot
Käyttötilanteisiin, joissa useamman henkilön pitää pystyä käyttämään samaa laitetta yhtäaikaan (koosta ja tekniikasta riippuen)
Pimeään, tekniikasta riippuen
Kirkkaaseen valoon, tekniikasta riippuen

<b>Graafinen ei sovi</b>
Käyttötilanteisiin, joissa käsiä käytetään muuhun tarkoitukseen
Käyttötilanteisiin, joissa silmiä käytetään muuhun tarkoitukseen
Käyttötilanteisiin, joissa vuorovaikutuslaite on käsien ulottumattomissa
Käyttötilanteisiin, joissa vuorovaikutuslaite on katseen ulottumattomissa
Käyttötilanteisiin, joissa komennon kohde on piilossa
Käyttötilanteisiin, joissa komennon kohde ei ole sillä hetkellä näkyvässä
Käyttötilanteisiin, joissa useamman henkilön pitää pystyä käyttämään samaa laitetta yhtäaikaan (koosta ja tekniikasta riippuen)
Pimeään, tekniikasta riippuen
Kirkkaaseen valoon, tekniikasta riippuen

<b>Puhe sopii</b>
Liikkuviin käyttötilanteisiin
Käyttötilanteisiin, joissa käsiä käytetään muuhun tarkoitukseen
Käyttötilanteisiin, joissa silmiä käytetään muuhun tarkoitukseen
Käyttötilanteisiin, joissa vuorovaikutuslaite on käsien ulottumattomissa
Käyttötilanteisiin, joissa vuorovaikutuslaite on katseen ulottumattomissa
Käyttötilanteisiin, joissa komennon kohde on piilossa
Käyttötilanteisiin, joissa komennon kohde ei ole sillä hetkellä näkyvissä
Käyttötilanteisiin, joissa useamman henkilön pitää pystyä käyttämään samaa laitetta yhtäaikaan (teknisin rajoituksin)
Kirkkaaseen valoon ja pimeään

<b>Puhe ei sovi</b>
Meluisiin ja kohinaisiin käyttötilanteisiin
Käyttötilanteisiin, joissa muut voivat häiriintyä äänestä
Käyttötilanteisiin, joissa käyttäjä haluaa säilyttää yksityisyytensä
Käyttötilanteisiin, joissa jotain sovelluksen tietoa pitää pystyä tarkastelemaan ajasta riippumatta
Käyttötilanteisiin, joissa viestistä halutaan peittää luonnolliset reaktiot

## 6.5.2 Vastaanottomodaliteetit

<b>Graafinen sopii</b>
Liikkuviin käyttötilanteisiin (riippuen koosta ja syöttölaitteesta)
Meluisiin ja kohinaisiin käyttötilanteisiin
Käyttötilanteisiin, joissa muut voivat häiriintyä äänestä
Käyttötilanteisiin, joissa käyttäjä haluaa säilyttää yksityisyytensä
Käyttötilanteisiin, joissa jotain sovelluksen tietoa pitää pystyä tarkastelemaan ajasta riippumatta
Käyttötilanteisiin, joissa viestistä halutaan peittää luonnolliset reaktiot
Jos käyttäjän kuuloaisti on jo ylikuormittunut muusta informaatiosta

<b>Graafinen ei sovi</b>
Käyttötilanteisiin, joissa silmiä käytetään muuhun tarkoitukseen
Käyttötilanteisiin, joissa vuorovaikutuslaite on katseen ulottumattomissa
Käyttötilanteisiin, joissa komennon kohde on piilossa
Käyttötilanteisiin, joissa komennon kohde ei ole sillä hetkellä näkyvissä
Käyttötilanteisiin, joissa useamman henkilön pitää pystyä käyttämään samaa laitetta yhtäaikaan (teknisin rajoituksin)
Jos käyttäjän näköaisti on jo ylikuormittunut muusta informaatiosta

<b>Puhe sopii</b>
Liikkuviin käyttötilanteisiin
Käyttötilanteisiin, joissa silmiä käytetään muuhun tarkoitukseen
Käyttötilanteisiin, joissa vuorovaikutuslaite on katseen ulottumattomissa
Käyttötilanteisiin, joissa useamman henkilön pitää pystyä kuulemaan samaa puolaute
Jos käyttäjän näköaisti on jo ylikuormittunut muusta informaatiosta

<b>Puhe ei sovi</b>
Meluisiin ja kohinaisiin käyttötilanteisiin
Käyttötilanteisiin, joissa muut voivat häiriintyä äänestä
Käyttötilanteisiin, joissa käyttäjä haluaa säilyttää yksityisyytensä
Käyttötilanteisiin, joissa jotain sovelluksen tietoa pitää pystyä tarkastelemaan ajasta riippumatta
Jos käyttäjän kuuloaisti on jo ylikuormittunut muusta informaatiosta

### 6.5.3 Paikka- ja kontekstiriippuvuus

Myös käyttötilanteiden suhteen graafinen ja puhekäyttöliittymä tukevat hyvin toisiaan. Kaikki modaliteetit eivät ole yhtä soveliaita kaikissa tilanteissa, joten soveltumattomat modaliteetit on pystyttävä sulkemaan käyttäjän tai järjestelmän toimesta.

Eryteisesti järjestelmän toimesta tapahtuva modaliteetin valinta on monimutkainen tehtävä. Järjestelmän täytyisi osata päätellä tai op-

pia käyttäjän yksilökohtaiset toiveet ja tulkita ne suhteessa käyttötilanteeseen eli se vaatii sekä käyttäjän että käyttöympäristön mallinnusta.

SmartKom Mobile on useassa eri ympäristössä toimiva henkilökohtainen digitaalinen avustaja. Peruskäyttötilanteet tapahtuvat jalkakulkijan ja autoilijan ympäristöissä. Taulukossa 6.1 on esitelty tavat, joilla käyttäjällä ja järjestelmällä on mahdollisuus mukauttaa modaaliteettivalikoimaa käyttötilanteeseen sopivaksi. (Bühler ym. 2002)

*Taulukko 6.1 SmartKom Mobilen modaaliteetti vaihdokset (Bühler ym. 2002)*

<b>Signal</b>	<b>System Reaction</b>
<i>User-driven transitions:</i>	
Push-to-activate	Suspend/Resume operation
Command/button	Turn on/off speech
Command/button	Turn on/off graphics
<i>System-driven transitions:</i>	
Install device in the car	Turn off gesture input
Take device out of the car	Switch to default mode
Start car	Turn off graphical displays
Stop car	Turn on graphical displays
Detect emergency	Suspend operation
Misrecognition/noise	Turn off speech input

Jameson (2002) esittää, että monissa käyttötilanteissa käyttäjän mahdollisuus modaaliteetin valintaan on ongelmallinen, koska itse valintatehtävä vaatii ylimääräistä kognitiivista toimintaa. Jameson kannattaakin ajatusta, että järjestelmä tutkisi passiivisesti (vrt. aktiivinen-passiivinen-sekoitus kappaleessa 3.6.1) käyttäjän ja ympäristön tilaa ja tekisi modaaliteetin valintaan liittyviä toimenpiteitä automaattisesti. Tällainen tarkkailu lähenee Rosalind Picardin (2000) lanseeraamaa tunnekoneiden (affective computing) käsitettä.

#### **6.5.4 Multimodaalisuus ja turvallisuus**

Osa modaliteeteista sitoo vahvasti käyttäjän tietyn aistin käytön aikana. Kaikissa käyttötilanteissa aistin varaaminen järjestelmän käyttöön ei ole turvallista. Tällaiset käyttötilanteet täytyy ottaa huomioon suunnitteluvaiheessa. Esimerkiksi autokäyttöliittymissä ei kuljettajan katsetta eikä käsiä saa sitoa muihin tehtäviin ajon aikana.

CarMMI autokäyttöliittymässä on pyritty siihen, että laitteen näytössä olisi mahdollisimman vähän huomiota herättävää liikettä ajon aikana (Neuss 2002). Edellisen kohdan SmartKom Mobilessa on menety askelen pidemmälle. Laitteesta sammutetaan eletunnistin, kun se asetetaan autotelineeseen, näyttö, kun auto käynnistetään ja koko laitteen toiminta keskeytetään, jos järjestelmä huomaa hätätilanteen (ks. taulukko 6.1).

Bernsen ja Dybkjær (2002) saivat ajosimulaattorin avulla suoritetussa testissä tulokseksi, että käyttäjät eivät katsoneet ajaessaan kovin usein autotietokoneen näyttöä joko turvallisuussyistä tai he eivät vain kokeneet sitä tarpeelliseksi. Tutkimuksessa kävi myös ilmi, että ruudulla olevan tiedon määrä on hyvä pitää keskitasolla. Liian vähän tietoa jää huomaamatta ja liian paljon saattaa viedä huomiota liiaksi.

#### **6.6 Modaliteetin valinta sisällön ja tehtävän perusteella**

Vuorovaikutukseen ja informaation esitystapaan liittyviin kysymyksiin on vaikea esittää tarkkoja vastauksia, koska ne täytyy ratkaista loppujen lopuksi sovelluskohtaisesti. Aiheeseen liittyen Oviatt (1999) huomauttaa, että kaikki modaliteetit eivät kykene välittämään samaa informaatiota.

Taulukossa 6.2. on esitetty muutamia yleisiä suosituksia graafisen käyttöliittymän (tässä tapauksessa suorakäyttö/direct manipulation käyttöliittymän) ja puhe käyttöliittymän sovellusalueista.

*Taulukko 6.2. Suorakäytön ja puheen ehdotetut sovellusalueet (Grasso ym. 1998).*

<b>Direct Manipulation</b>	<b>Speech Recognition</b>
Visible References	Nonvisible References
Limited References	Multiple References
Simple Actions	Complex Actions

Sisältöön ja tehtävään liittyviä suosituksia ovat lähinnä toimenpiteiden monimutkaisuuteen ja kohteiden viittausmahdollisuuksiin liittyvät sovellukset.

Monet kappaleessa 5. esitellyt modaaliteettien ominaisuudet vaikuttavat käyttäjän ja käyttötilanteen lisäksi myös sisältöön. Taulukkoon 6.3 on poimittu kappaleessa 5.7 esitettyjen 25:n modaaliteettiominaisuuden joukosta sisältöön ja tehtävään eniten liittyvät ominaisuudet.

*Taulukko 6.3 Sisältöön ja tehtävään liittyvät modaaliteettiominaisuudet*

Linguistic input/output modalities have interpretational scope, which makes them eminently suited for conveying abstract information. They are therefore unsuited for conveying high-specificity information including detailed information on spatial manipulation and location.
Linguistic input/output modalities, being unsuited for specifying detailed information on spatial manipulation, lack an adequate vocabulary for describing the manipulations.
Static graphic/haptic input/output modalities allow the simultaneous representation of large amounts of information for free visual/tactile inspection and subsequent interaction.
Dynamic input/output modalities, being temporal (serial and transient), do not offer the cognitive advantages (wrt. Attention and memory) of freedom of perceptual inspection.
Speech input/output modalities, being temporal (serial and transient) and non-spatial, should be presented sequentially rather than in parallel.
Non-spontaneous speech input modalities (isolated words, connected words) are unnatural and add cognitive processing load.
Discourse input/output modalities have strong rhetorical potential.
Discourse input/output modalities are situation-dependent.
Spontaneous spoken labels/keywords and discourse input/ output modalities are natural for



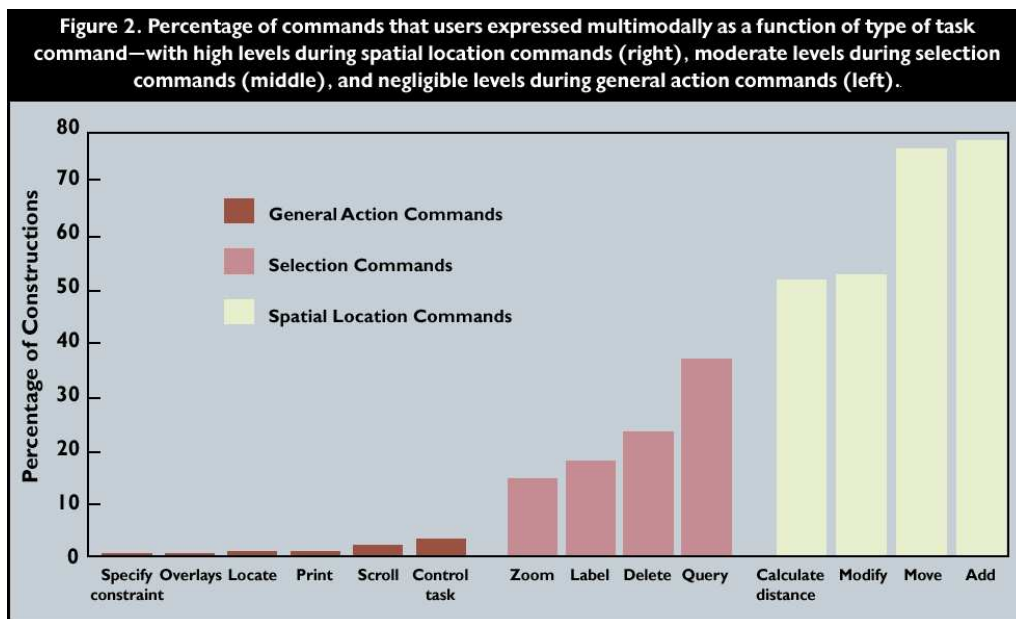
humans in the sense that they are learnt from early on (by most people and in a particular tongue and, possibly, accent). (Note that spontaneous keywords and discourse must be distinguished from designer-designed keywords and discourse, which are not necessarily natural to the actual users.)
Direct manipulation selection input into graphic output space can be lengthy if the user is dealing with deep hierarchies, extended series of links, or the setting of a large number of parameters.
Haptic deictic input gesture is eminently suited for spatial manipulation and indication of spatial location. It is not suited for conveying abstract information.
Linguistic text and discourse input/output modalities have very high expressiveness.
Images have specificity and are eminently suited for representing high-specificity information on spatio-temporal objects and situations. They are therefore unsuited for conveying abstract information.
Speech input/output modalities, being physically realised in the acoustic medium, possess a broad range of acoustic information channels for the natural expression of information.

### 6.6.1 Tuottomodaliteetit

Edellä esiteltyjen ominaisuuksien lisäksi seuraavassa tuodaan esille muutamia kokeellisesti havaittuja ilmiöitä tuottomodaliteetin suhteesta tehtävään ja sisältöön.

Ihmiset eivät välttämättä käytä multimodaalisia sovelluksia kaikissa tehtävissä multimodaalisesti, jos järjestelmä antaa heidän valita. Käyttö voi olla vuorotellen yksi- ja multimodaalista tehtävästä riippuen. Kuvassa 6.3. esitetty pylväsdiagrammi kuvaa käyttäjien mielipidettä siitä, kuinka multimodaalisia komentoja kynä- ja puheohjauksella varustetussa karttasovelluksessa. (Oviatt ym. 1997)

Kuvasta 6.3. voidaan huomata, että erityyppisissä tehtäväryhmissä multimodaalisten komentojen käytön määrä vaihtelee huomattavasti, vaikka kaikki tehtävät olisi ollut mahdollista suorittaa multimodaalisesti. Erityisesti käyttäjät suosivat multimodaalista käyttöä tilallisia viitteitä tarvitsevien komentojen yhteydessä (Oviatt 1996, Oviatt ym. 1997).



Kuva 6.3. Multimodaalisten komentojen osuus tehtävätyypeittäin (Oviatt 1999).

Aikaisemmissa tutkimuksissa kynä- ja puheohjatun graafisen sovelluksen kanssa Oviatt ja Olsen (1994) olivat havainneet, että ihmisillä on tapana käyttää eri modaliteetteja toisiaan täydentäen. Alkuperäinen ja korjaustieto välitettiin eri modaliteettien kautta, samoin data ja komennot, luvut ja muu teksti sekä luvut ja niihin viittaavat kuvaukset. Modaliteetit eivät olleet sidottuja tiettyyn toimintaan, vaan esimerkiksi alkuperäinen tieto välitettiin puheena ja korjaus kirjoittamalla yhtä usein kuin alkuperäinen kirjoittamalla ja korjaus puheena. Muita huomiota oli, että luvut ja erisnimet kirjoitettiin useammin kuin muu tekstipohjainen sisältö. Samoin tiedon esitysmuodolla oli vaikutusta niin, että lomake-esitys nosti merkittävästi kynän käyttöä tiedon syöttöön. Karttasovellusesimerkissä subjekti, verbi ja objekti syötettiin yleensä puheella ja paikkaan liittyvä informaatio kirjoittamalla (Oviatt ym. 1997, Oviatt 1999).

Grasso ym. (1998) tutkivat multimodaalista tiedonsyöttöä havaintorakenneteorian (theory of perceptual structure) pohjalta. Havaintorakenneteorian mukaan rakenteen ulottuvuudet ovat yhteneväisiä (integral), jos niihin ei erikseen voida kiinnittää huomiota. Muussa

tapauksessa ulottuvuudet ovat erotettavissa olevia (separable). Aiemmissa kokeissa oli huomattu, että syöttölaite sopii paremmin tehtävään, jos sen ulottuvuudet vastaavat tehtävän havaintorakenteen ulottuvuuksia. Grasso kumppaneineen selvitti, miten havaintorakenteen ulottuvuudet suhteutuvat modaliteettien valintaan. Lopputulos oli, että tehtävän suoritus aika lyheni, tarkkuus parani ja käyttäjäytyytyväisyys kasvoi, kun tehtävän yhteneväiset ominaisuudet syötettiin yhden modaliteetin kautta. Hypoteesia, jonka mukaan erotettavissa olevien ominaisuuksien syöttäminen toimisi paremmin eri modaliteettien kautta, ei voitu varmistaa, koska tarvittava koejärjestely ei kuulunut tutkimukseen. Tehtävän havaintorakenne kannattaa kuitenkin ottaa huomioon silloin, kun sisällön jakamista eri modaliteetteihin suunnitellaan.

#### **6.6.2 Vastaanottomodaliteetit**

Alistair Sutcliffe (2000) on tutkinut ja analysoinut medioiden valintaa ja niiden yhdistämistä multimedialjärjestelmien yhteydessä. Tässä tapauksessa media tarkoittaa tietynlaista tiedon esitystapaa. Ylimmällä tasolla mediaelementit jakautuvat luonnollisiin ja suunniteltuihin, jotka molemmat jakautuvat kuva- ja äänielementteihin, jotka puolestaan jakautuvat staattisiin ja dynaamisiin medioihin. Lisäksi oma haaransa on kieleen pohjautuvat mediaelementit, jotka kuuluvat suunniteltujen luokkaan ja jakautuvat edelleen kuva- ja äänielementteihin. Lisäksi hän jakaa tiedon 12 erilaiseen tyyppiin. Edellä kuvattuun luokitteluun perustuen Sutcliffe esittelee esimerkin mediaelementtien valinnasta (taulukko 6.4).

Taulukko 6.4 Mediaelementtien valinta (Sutcliffe 2000).

Informaation tyyppi	Ensisijainen media	Esimerkki
Fyysinen (aineellinen)	Pysähtynyt tai liikkuva kuva	Rakennuskaavio
Abstrakti (käsitteellinen)	Teksti tai puhe	Myyntiehtojen selitys
Kuvaileva	Teksti tai puhe	Kemialliset ominaisuudet
Tilallinen	Realistinen media / valokuva	Henkilön kasvot
Lukuarvo	Teksti/taulukot/numeeriset listat	Paine lukemat
Suhteet (arvot)	Suunnitellut kuvat/graafit, kaaviot	Pylväsdiagrammi kuukausittaisesta sademäärästä
Menettelytapoihin liittyvä	Kuvasarjat, teksti	Evakuointiohjeet
Erillinen liike	Pysähtynyt kuva	Kahvin keitto
Jatkuva liike	Liikkuva kuva	Liikkeet hiihdon aikana
Tapahtumat	Ääni, puhe	Palohälytys
Tilat	Pysähtynyt kuva, teksti	Kuva sääoloista
Syy-yhteys	Liikkuva kuva, teksti, puhe	Video sademyrskystä, joka aiheuttaa äkillisen tulvan

Taulukko 6.4 on karkea yleisen tason suositus, mutta siitä voidaan kuitenkin nähdä, että joku media voi olla parempi kuin joku toinen tietyyppisen informaation esitykseen. Jo aikaisemminkin ilmi tullut asia toistuu tässäkin tapauksessa eli graafinen modality sopii tilallisiin ja tarkkoihin asioihin ja puhe taas puolestaan abstrakteihin ja kuvaileviin asioihin. Taulukkolaskenta on helpompaa graafisesti ja omenan maku on helpompi kuvailla puheena kuin esittää kuvana. Usein on tarpeellista yhdistää useita eri medioita ajallisessa ja/tai tilallisessa ulottuvuudessa, kun esimerkiksi pyritään esittämään menettelytapoja tai syy-yhteyksiä (Sutcliffe 2000).

Norman (1993: 107) analysoi visuaalis-haptisen ja auditiivisen vuorovaikutuksen eroja esimerkissään lentoaikataulupalveluista ja esittää samalla, kuinka tehtävää ja sisältöä voidaan lähestyä myös af-

fordanssien kautta. Lentoyhtiöllä oli kaksi vaihtoehtoista palvelua lentoaikataulujen tarkastamiseksi. Toinen oli perinteinen pieni mukana kuljetettava aikataulukirja ja toinen puhelimen näppäimillä käytettävä automaattinen puhepalvelujärjestelmä. Käytännössä puhepalvelujärjestelmä oli yli kolme kertaa hitaampi halutun lennon ajan selvittämisessä. Suurimmat ongelmat olivat puheen hetkellisyys eli tiedon pysymättömyys, rinnakkaisen tiedon esittämisen mahdottomuus ja hitaus vastaanottomodaliteettina. Hetkellisyys ja rinnakkaisuuden puute aiheuttivat sen, että eri lentojen aikoja oli hankala verrata keskenään ja toisaalta ne edellyttivät monitasoista hierarkkista navigointirakennetta puhekäyttöliittymään. Eräs ongelma oli myös korjausmahdollisuuksien puute tilanteessa, jossa käyttäjä painoi vahingossa väärää näppäintä. Tässä tapauksessa kyseessä oli enemmän järjestelmän ongelma, mutta yleensä ottaen virheenkorjaus on eräs keskeisiä asioita puhekäyttöliittymissä (Shum ym. 2001).

Eltling ym. (2002) tutkivat tekstin, puheen ja kuvien erilaisten yhdistelmien havainnointia erilaisissa päätelaitteissa (PC, PDA, TV) turistiopassovelluksella. Lopputuloksena oli, että käytetty laite vaikutti jonkun verran tuloksiin, kuitenkin lähinnä niin, että PDA:n hankala käsiteltävyys heikensi testituloksia. Modaliteettiyhdistelmistä käyttäjät pitivät eniten tekstin, kuvan ja puheen yhdistelmästä. Muistamistestissä ykköseksi nousi puheen ja kuvan yhdistelmä niin kuin oli ennustettukin. Puhe ja kuva käyttävät eri työmuistin prosesseja ja aiheuttivat siten vähemmän kognitiivista kuormaa, mikä puolestaan jättää enemmän kapasiteettia muistinkäsittelyä varten. Tekstin lukeminen lisää kognitiivista kuormaa ratkaisevasti.

### **6.6.3 Tehtävän sisäinen tilanneriippuvuus**

Monet järjestelmän avulla tehtävistä asioista voidaan jakaa osatehtäviin. Toiset modaliteetit saattavat soveltua eri osatehtäviin toisia

paremmin. Modaliteettien valinta osatehtäviin riippuu käyttäjästä ja käyttötilanteesta.

Esimerkiksi CarMMI autokäyttöliittymässä käyttäjät vaihtoivat modaliteettia 0,8 kertaa tehtävän aikana lopullisissa käyttäjätesteissä. Muutenkin testihenkilöt käyttivät eri modaliteetteja ahkerasti hyväkseen. (Neuss 2002)

## **6.7 Muuta yhdistämisessä huomioitavaa**

### **6.7.1 Konsistenssi modaliteettien välillä**

Eräs käyttöliittymäsuunnittelun perussääntöjä on yhtenäisyyden säilyttäminen koko käyttöliittymässä. Jos käyttöliittymään kuuluu useita modaliteetteja, yhtenäisyyden ylläpitämisestä tulee mielenkiintoinen tehtävä.

Kirjallisuudesta ei löydy suoria esimerkkejä yhtenäisyyden ylläpitämisestä, mutta käytännössä ainakin termit kannattaa yrittää yhtenäistää kummassakin modaliteetissa. Lisäksi saattaa olla mahdollista suunnitella valikkorakenteet saman kaltaisiksi sekä graafisessa että puhekkäyttöliittymässä.

### **6.7.2 Painotus (visuaalis- vai puhepainotteinen)**

Modaliteetteja voidaan painottaa suhteessa toisiinsa. Joku modaliteeteista voi olla määräävämmässä asemassa ja muut toimia avustavassa roolissa.

Painotuksen harkinta saattaa tulla käytännössä eteen, kun olemassa olevia sovelluksia aletaan muuttamaan multimodaalisiksi. Silloin on yleensä järkevää antaa alkuperäisen sovelluksen olla päämodaliteetti ja uusien toimia apumodaliteetteina.

Visuaalispainotteiseen käyttöliittymään suositellaan lyhyempiä puhepalautteita verrattuna puhepainotteiseen, koska visuaalispainotteisessa suurempi osa informaatiosta tulee visuaalisen kanavan kautta.

## 7. Ihmisen ja koneen kohtaaminen

### 7.1 *Epäsymmetrinen käyttötilanne*

Suchman (1987) väittää, että ihminen toimii vahvasti tilanteen mukaan, sen sijaan, että erilaiset suunnitelmat määräisivät toimintaa. Suchmanin mukaan suunnitelmat toimivat toimintaa ohjaavina resursseina, mutta käytännössä itse toiminta tapahtuu aina suhteessa tilanteeseen. Ihminen reagoi tilanteeseen tarpeen mukaan ja muokkaa suunnitelmaa tarpeen vaatiessa.

Ihmisen ja koneen vuorovaikutuksen tekee ongelmalliseksi niiden epäsymmetrisyys vuorovaikutustilanteen suhteen (Suchman 1987). Kone aistii tilanteesta sen, mitä vajavaisilta sensoreilta tulevasta tietomäärästä voidaan prosessoida. Prosessointi on myös yleensä suhteellisen yksinkertaista ja vaikuttaa vain tiettyihin kohtiin koneen toimintaprosessissa eli suunnittelijan laatimassa koneen toimintasuunnitelmassa (vrt. kappale 4.3 Norman). Ihmisellä on paljon laajemmat mahdollisuudet aistia tilanteen tapahtumia ja muodostaa niistä kokonaisvaltaisia tulkintoja.

Toinen asia, joka olennaisesti lisää koneen ja ihmisen välistä epäsymmetriaa, on ihmisten taustatietomäärä. Jokaisella ihmisellä on yksilöllisesti linkittynyt valtava tietomäärä maailman rakenteesta, ihmisten välisistä kommunikaatiotavoista, päivän säästä ja monista muista asioista. Kaikkialla samojen fysiikan lakien, yhteisen kulttuurin ja monien muiden asioiden tukemana ihmisellä on tunne, että keskustelukumppanilla on vastaavat tiedot asioista. Ihminen voi viitata yhteisiin asioihin ja olettaa asioita, joita ei ole sanottu ääneen, kunnes vastapuoli osoittaa merkkejä siitä, ettei hän ymmärrä. Sen lisäksi, että taustatietomäärä on huomattava, on sitä vaikea mallintaa (Minsky 2000).



Lisäksi asia, joka olennaisesti lisää koneen ja ihmisen välistä epäsymmetriaa on ihmisten "luontainen" kyky korjata keskustelussa ilmenevät väärinymmärrykset. Normaalissa keskustelussa esiintyy jatkuvasti keskusteluvuoroja, joiden tarkoitus on tarkistaa, että vastapuoli on ymmärtänyt, mitä puhuja on yrittänyt sanoa ja vastavasti korjata mahdolliset väärinkäsitykset. Ihmisillä on siis hyvät työkalut yhteisymmärryksen rakentamiseen. Sovelluksen suunnittelijan täytyy sitä vastoin ottaa huomioon mahdolliset virhetilanteet jo etukäteen.

Tietojenkäsittelyteorioiden edistymisen, laskentatehon lisääntymisen ja sensoritekniiikan paranemisen myötä epäsymmetrisyys edellisten ominaisuuksien suhteen on pienentynyt ja pienenee yhä. Multimo-  
daaliset käyttöliittymät ovat yksi tapa pienentää epäsymmetrisyyttä.

## **7.2 Työkalu vs. toimija**

HCI-keskustelussa on kaksi toisilleen ainakin osittain vastakkaista näkemystä tietokoneen roolista vuorovaikutustilanteessa. Perinteinen tapa on nähdä tietokone työkaluna (computer as tool), joka tekee, mitä käyttäjä käskää. Toinen tapa on nähdä tietokone toimijana (computer as agent), joka lähtee ajatuksesta, että tietokone on jollain tasolla aktiivinen toimija vuorovaikutustilanteessa, eikä ainoastaan passiivinen työkalu.

Vastakkainasetteluun liittyvät olennaisesti hallinnan ja luottamuksen käsitteet. Työkaluajattelun mukaan käyttäjällä täytyy olla ainakin näennäinen tunne työkalunsa hallinnasta, jotta sen käyttö voisi olla tyydyttävää. Käyttäjän luottamus kohdistuu lähinnä häneen itseensä, siihen, että hän osaa käyttää työkalua oikein. Toimijaajattelussa käyttäjä luovuttaa osan hallinnasta tietokoneelle, jolloin käyttäjän täytyy luottaa, että kone toimii hänen toiveidensa mukaan. Sekä hallinta että luottamus ovat käytännössä hyvin subjektiivisia asioita (Karat, Karat & Ukelson 2000).

Työkaluajattelun ongelma on siinä, että käyttäjät joka tapauksessa liittävätkin tietokoneeseen helposti toimijuuteen viittaavia piirteitä. Tietokonetta voidaan esimerkiksi säätää tai maanitella tai ääritapauksissa lyödä, jos se ei suostu tekemään, mitä käyttäjä tahtoo. Toiset saattavat ihmetellä, jos tietokone tekee omin päin jotain näennäisen järkevää.

Lapset saattavat pitää tietokonepohjaisia artefakteja ”melkein elävinä”. He saattavat liittää elollisuuden objekteihin, jotka reagoivat ympäristön tapahtumiin tai liikkuvat itsestään. Toisaalta, jotta lapset kokisivat jonkun kokonaisuuden inhimilliseksi, täytyy sen osoittaa tunteita, selkeää ajattelua, puhua tai toimia määrätietoisesti. (Turkile 1984. Suchmanin 1987:5 mukaan)

Lucy Suchman (1987) esittää, että piirteet, jotka tekevät tietokoneesta vuorovaikutteisen ihmisten mielissä, ovat reaktiivisuus, kielellisyys ja läpinäkymättömyys. Reaktiivisuus tarkoittaa välitöntä palautetta käyttäjän toimista. Kielellisyys tarkoittaa sitä, että tietokonetta käytetään suoravaikutteisten vipujen ja nappien sijaan kielellisten operaatioiden (esim. ”järjestä tiedostot”) kautta. Läpinäkymättömyys tarkoittaa sitä, että käyttäjä ei voi järjestelmän monimutkaisuudesta johtuen tietää, mitä koneen sisällä tapahtuu, samoin kuin hän ei voi tietää, mitä toisen ihmisen päällä liikkuu heidän keskustellessaan. Läpinäkymättömyydestä seuraa, että tietokoneen toimintaan liittyy mahdollisuus odottamattomiin tilanteisiin. Samat piirteet vaikuttavat myös siihen, että tietokonetta pidetään helposti jonkin tason toimijana työkalun sijaan.

Jos laite osoittaa jollain tavalla inhimillistä käytöstä, ihmisillä on tapana päätellä sen perusteella, että laiteella saattaa olla muitakin inhimillisiä ominaisuuksia (Nass & Gong 2000, Suchman 2000). Työkaluajattelun kannattajat vastustavatkin antropomorfista (inhimillis-

tävää) ja animistista (elollistavaa) suunnittelua, jottei käyttäjälle tulisi vääriä mielikuvia sovelluksen toiminnasta ja sitä kautta ongelmatilanteita ja ahdistusta silloin, kun käyttäjän malli ei vastaisikaan järjestelmän mallia. (Shneiderman 1998).

Sama ihmisten tapa kuvitella havaittua enemmän inhimillisiä ominaisuuksia on keskeinen ilmiö myös toimija-ajattelussa. Siinä tapauksessa kysymys on kuitenkin ennemminkin haasteesta kuin ongelmasta eli miten suunnitella käyttöliittymä niin, että se mahdollisimman hyvin vastaisi ihmisen toimijuudesta saamaa mielikuvaa.

### **7.3 Puheen erityisasema**

Ihmisen on vaikea muuttaa automatisoituneita kokemukselliseen toimintaan liittyviä prosesseja. Niistä on opittava pois ennen kuin niitä pystyy muuttamaan (Sinkkonen ym. 2002). Voidaan ajatella, että mitä inhimillisempi vuorovaikutustapa on, sitä automatisoituneempi se on ihmisillä. Monet multimodaalisissa sovelluksissa käytetyt vuorovaikutustavat ovat lähempänä ihmisten välistä vuorovaikutusta kuin perinteiset näyttö, hiiri ja näppäimistö, joten ilmiö on syytä ottaa huomioon suunnittelussa.

Erityisesti puhe, äänenpainoinen, ilmeinen ja eleinen on evoluution myötä mahdollistunut ja jo syntymästä alkaen opittu syvästi automatisoitunut kommunikointitapa. Nass ja Gong (2000) väittävät, ettei ihmisillä ole kykyä nujertaa puhumiseen liittyviä perusvaistoja ja että ihmiset käyttävät samoja sääntöjä myös keinokehoisten puhejärjestelmien toiminnan arviointiin.

Väitteidensä tueksi Nass ja Gong esittävät mielenkiintoisia tutkimustuloksia puheen evolutionäärisiin ominaisuuksiin perustuvista koikeista. Ihmisillä on tapana alkaa hyperartikuloimaan eli puhumaan liioitellun selkeästi vieraskieliselle tai huonokuuloisille, jos nämä eivät tunnu ymmärtävän tavallista puhetta, jotta viesti menisi perille.

Testikäyttäjät alkavat hyperartikuloimaan samalla tavoin puheen-tunnistusjärjestelmälle, jos se ei näytä ymmärtävän heitä. Testikäyttäjät antoivat myös huomattavasti sosiaalisesti hyväksyttävämpiä vastauksia, jos he joutuivat täyttämään kyselyn puheella verrattuna saman kyselyn täyttämiseen hiirellä, näppäimistöllä tai käsin kirjoittamalla. Tulos tuki hypoteesia, jonka mukaan puhe edustaa inhimillisyyttä ja sen kautta sosiaalista läheisyyttä.

Ihmiset olivat erittäin herkkiä arvioimaan sukupuolta äänen perusteella. Testikäyttäjät liittivät tietokoneeseen sukupuoleen liittyviä ominaisuuksia riippuen siitä käytettiinkö saman sisällön kertomiseen mies- vai naisääntä. Vastaavasti käyttäjät tunnistivat sukupuoli- ja persoonallisuustyyppisiä sekä tunnesisältöjä selkeästi syntetisoidusta äänestä. Käyttäjät pitivät myös eri ääniä eri persoonalluuksina vaikka ne olisivat tulleet fyysisesti samasta koneesta. Tietokone-toimijan käytöksestä annettiin johdonmukaisesti positiivisempia arvioita, jos kyselyn teki sama toimijan ääni verrattuna tilanteeseen, jossa kyselyn teki toinen ääni. Samoin testikäyttäjät kokivat eron objektiivisuuden ja itsekehun välillä kun eri ääni kehuu tietokone-toimijan käytöstä verrattuna tilanteeseen, jossa toimijan oma ääni olisi suorittanut kehumisen.

Nass ja Gong esittelivät myös multimodaalisuuden kannalta mielenkiintoisia havaintoja, joiden mukaan kasvot ja puhe tunnistetaan kokonaisvaltaisesti. Tunnetun McGurkin ilmiön mukaan sama ääni tunnistetaan eri äänneiksi huulten asennosta riippuen. Toisen kokeen perusteella visuaalisen ja puhemodaliteetin yhtenäisyys havaitaan. Testissä koehenkilöille esitettiin avoimia ja kyllä/ei-tyyppisiä henkilökohtaisia kysymyksiä syntetisoidulla ja etukäteen äänitetyllä puheella sekä graafisen keinopään kanssa että ilman. Pelkkä nauhoitettu ääni ja syntetisoitu ääni syntetisoidun pään kanssa saivat aikaan merkittävästi enemmän ja syvempää keskustelua kuin pelkkä syntetisoitu puhe tai nauhoitettu puhe syntetisoidun pään kanssa.

Multimodaalisten sovellusten kohdalla sovelluksen toimijuus muuttuu, kun muut modaliteetit tulevat mukaan vuorovaikutustilanteeseen. Oviatt (1999) on esimerkiksi havainnut, että käyttäjät puhuvat normaalia lyhyempiä ja yksinkertaisempia lauseita multimodaalisesti kommunikoidessaan, koska osa informaatiosta välittyy muiden modaliteettien kautta.

## 8. Loppupäätelmät

Tämän työn myötä olen päässyt tutustumaan pintaa syvemältä multimodaalisten käyttöliittymien maailmaan. Multimodaalisuus näyttää sisältävän huimia mahdollisuuksia, mutta myös runsaasti haasteita.

Multimodaalisuus tuntuu tarjoavan parannusta juuri niihin ongelmiin, joita käyttöliittymäsuunnittelijat kohtaavat päivittäin. Multimodaalisuuden lupaukset parantaa erilaisten ihmisten mahdollisuuksia käyttää nykyaikaisia tietokonesovelluksia ja mukautua vaihteleviin käyttötilanteisiin kuulostavat todella hyviltä. Tuntuu kuitenkin siltä, että käytännön sovellusten yleistymiseen on vielä pitkä matka.

Uusien modaliteettien mukaantulo laajentaa suunnitteluavaruutta entisestään ja lisää samalla tehtävien valintojen määrää. Alan tutkimus on tällä hetkellä siinä kehitysvaiheessa, että se ei pysty vielä tarjoamaan kattavia neuvoja siitä, miten modaliteetteja tulisi käyttää ja miten ne pitäisi yhdistää. Ihmisen kognitiivisia prosesseja ei tunneta vielä tarpeeksi hyvin niin, että pystyttäisiin ennustamaan, mitä tapahtuu, kun havainnot ja toiminta ylittävät modaliteettien rajat. --Empiiristä tietoakaan ei ole vielä kertynyt tarpeeksi käytännön sovellusten ja käyttäjätutkimusten kautta, jotta sen perusteella voitaisiin tehdä kattavia johtopäätöksiä.

Joiltain sovellusalueilta alkaa pikkuhiljaa kertymään käytännön tietoaakin. Esimerkiksi Sarah Oviattin tutkimukset puhe- ja kynäkäyttöisten käyttöliittymien parissa ovat nostaneet esille monia mielenkiintoisia ilmiöitä. Käytännön työvälineiden puuttuessa kokemuspäisen tiedon kertyminen on ollut toistaiseksi hidasta. Toivottavasti tilanne paranee tulevaisuudessa.

Tuntuu siltä, että tähän saakka merkittävä osa tutkimuksesta on ollut varsin teknologiavetoista, mikä on hyvin ymmärrettävää, koska multimodaalisuus tarjoaa paljon mielenkiintoisia haasteita sillä alueella. Usein multimodaalisuutta perustellaan esimerkiksi puheen tai eleiden tunnistustarkkuuden paranemisella, mikä on hyvä asia, koska se hyödyttää käyttäjää, mutta loppujen lopuksi se ei liene kuitenkaan tärkein syy käyttää multimodaalista sovellusta.

Onko kukaan tutkinut haluavatko ihmiset ylipäättään käyttää autordiota puhumalla tai viittomalla, kun se on tähänkin asti onnistunut nappeja pyörittämällä? Monet erityisryhmiin kuuluvat ihmiset sanoisivat varmaan, että he kyllä haluavat, koska tähän mennessä laitteiden käyttäminen on ollut niin vaikeaa.

Graafisen ja puhekäyttöliittymän yhdistelmä, mahdollisesti osoitinlaitteella höystettynä, tuntuu järkevältä ratkaisulta, koska modaliteetit täydentävät toisiaan. Käyttäjän kannalta tuntuu fiksuimmalta yhdistää modaliteetit tasa-arvoisesti niin, että käyttäjä voi mahdollisuksiensa ja halujensa mukaan valita sopivan yhdistelmän ja vaihdella sitä tarpeen mukaan, kun käyttötilanteet vaihtuvat. Samalla myös erityisryhmien rajoitetut mahdollisuudet tulevat huomioon otetuksi.

Tässä työssä vähemmälle käsittelylle jääneestä järjestelmän näkökulmasta modaliteetit olisi järkevää yhdistää toisiaan vahvistavasti niin, että yhden modaliteetin heikkoja tunnistustuloksia voitaisiin varmentaa ja monitulkintaisia tuloksia tarkentaa toisista modaliteeteista saadulla informaatiolla. Ihmiset eivät tunnu kuitenkaan luonnostaan antavan ylimääräistä informaatiota aktiivisesti muiden modaliteettien kautta. Toimiva ratkaisu voisi olla ihmisen tiedostamattaan antaman ylimääräisen tiedon tarkkailu vaivihkaa järjestelmän toimesta. Muita haasteita järjestelmän suunnitteluun on erilaisten

ihmisten ja vaihtelevien käyttötilanteiden vaatima käyttäjän ja ympäristön mallintaminen.

Käyttäjäkokemuksen kannalta graafisen ja puhekäyttöliittymän yhdistelmä on mielenkiintoinen, koska siinä yhdistyvät selkeästi keino-tekoinen ja syvästi inhimillinen vuorovaikutuskanava keskenään. Ihmisten tapa liittää toimijuuteen viittaavia ominaisuuksia tietokoneen kaltaisiin laitteisiin on todellinen, eikä siitä aiheutuviin ongelmiin ole kyetty vielä kehittämään täydellistä ratkaisua. Toinen käyttäjän kannalta tärkeä asia on turvallisuus. Multimodaaliset käyttöliittymät saattavat vangita kaikki tilanteen kannalta tärkeät aistit, mikä saattaa esimerkiksi liikenteessä aiheuttaa ongelmia.

Olen kokenut työn tekemisen varsin opettavaisena sekä tiedollisesti että prosessina. Koen oppineeni paljon ja tunnen, että omaan nyt huomattavasti paremmat mahdollisuudet arvioida multimodaalisuuteen ja käyttöliittymiin liittyviä asioita. Työn tekeminen on ollut mielenkiintoista, koska aiheen ympärillä ihmistieteet ja tekniikka yhdistyvät entistä laajemmin. Jään odottamaan mielenkiinnolla, miten multimodaaliset sovellukset kehittyvät tulevaisuudessa.



## Lähdeluettelo

- Balentine, B. ja Morgan, D. P. (1999). *How to Build a Speech Recognition Application*. Enterprise Integration Group.
- Bernsen, N. O. (2002). Multimodality in Language and Speech Systems - From Theory to Design Support Tool. Granström, B., House, D. & Karlsson, I. (ed.) 2002. *Multimodality in Language and Speech Systems*. Text, Speech and Language Technology (Vol. 19). Kluwer Academic Publishers. [myös: <http://www.nis.sdu.dk/~nob/publications/MILASS-CHAP-3.10.pdf>, 7.4.2003].
- Bernsen, N. O. ja Dybkjær, L. (2002). *A Multimodal Virtual Co-driver's Problems with the Driver*. Proceedings of the ISCA Tutorial and Research Workshop on Multi-Modal Dialogue in Mobile Environments, June 17-19, 2002, Germany.
- Bohus, D. ja Rudnicky, A. (2002). *LARRI: A Language-based Maintenance and Repair Assistant*. Proceedings of the ISCA Tutorial and Research Workshop on Multi-Modal Dialogue in Mobile Environments, June 17-19, 2002, Germany.
- Bolt, R. A. (1980). "Put-That-There": Voice and Gesture at the Graphics Interface. *ACM Computer Graphics* 14, 3, 262-270.
- Boyce, S. (2000). *Natural Spoken Dialogue Systems for Telephony Applications*. *Communications of the ACM* September 2000 (Vol.43, No.9).
- Bühler, D., Minker, W., Häussler, J., ja Krüger S. (2002). *The Smartkom Mobile Multi-Modal Dialogue System*. Proceedings of the ISCA Tutorial and Research Workshop on Multi-Modal Dialogue in Mobile Environments, June 17-19, 2002, Germany.
- Gibbon, D., Mertins, I. ja Moore, R. (toim.) (2000). *Handbook of multimodal and spoken dialogue systems: resources, terminology, and product evaluation*. Kluwer Academic Publishers.
- Grasso, M. A., Ebert, D.S. ja Finin, T. W. (1998). *The integrality of speech in multimodal interfaces*. *ACM Transactions on Computer-Human Interaction* December 1998 (Vol.5, No.4).
- Dix, A., Finlay, J., Abowd, G. ja Beale, R. (1998). *Human-Computer Interaction (second edition)*. Prentice Hall Europe.

Elting, C., Zwickel, J. ja Malaka R. (2002). *Device-dependent modality selection for user-interfaces - an empirical study*. Proceedings of the 7th International Conference on Intelligent User Interfaces January 2002.

Gustafson, J., Bell, L., Boye, J. ja Wirén, M. (2002). *Constraint Manipulation and Visualization in a Multimodal Dialogue System*. Proceedings of the ISCA Tutorial and Research Workshop on Multi-Modal Dialogue in Mobile Environments, June 17-19, 2002, Germany.

Hemsen, H. (2002). *A Testbed for Evaluating Multimodal Dialogue Systems for Small Screen Devices*. Proceedings of the ISCA Tutorial and Research Workshop on Multi-Modal Dialogue in Mobile Environments, June 17-19, 2002, Germany.

HERO (2003). *Helsingin seudun erilaiset oppijat ry:n verkkosivut*. [<http://www.lukihero.fi/index2.html>, 15.4.2003].

Hollier, M. P., Rimell, A. N., Hands, D. S., ja Voelcker, R. M. (1999). *Multi-modal perception*. BT Technol J January 1999 (Vol.17, No.1).

Hurtig, T. (2002). *Puhuva kartta*. Esitys Interactin päätösdemossa 12.12.2002. Taideteollinen korkeakoulu. [Myös: <http://mlab.uiah.fi/interact/demot/demo12122002/2002.12.12.USIX-IA-loppudemo.ppt>, 20.4.2003]

ISO 13407 (1999). *Human centered design processes for interactive systems*. [Keskeiset asiat: <http://www.usabilitynet.org/tools/13407stds.htm>, 20.4.2003].

Jameson, A. (2002). *Usability Issues and Methods for Mobile Multimodal Systems*. Proceedings of the ISCA Tutorial and Research Workshop on Multi-Modal Dialogue in Mobile Environments, June 17-19, 2002, Germany.

Jokinen, K., Kerminen, A., Kaipainen, M., Jauhiainen, T., Wilcock, G., Turunen, M., Hakulinen, J., Kuusisto, J. ja Lagus, K. (2002). *Adaptive Dialogue Systems - Interaction with Interact*. 3rd SIGdial Workshop on Discourse and Dialogue, July 11-12, 2002, Philadelphia.

Jokinen, K. ja Raike, A. (2002). *Multimodality – the latest technology and visions and demands for the future*. Multimodality IT-seminar, Castberggård, Denmark.

Karat, J., Karat, C-M. ja Ukelson J. (2000). *Affordances, motivation, and the design of user interfaces*. Communications of the ACM August 2000 (Vol.43, No.8).

- Keinonen, T. (toim.) (2000). *Miten käytettävyys muotoillaan*. Taideteollinen korkeakoulu.
- Laurel, B. (toim.) (1990). *The Art of Human Computer Interface design*. Apple Computer.
- Lehtonen, J. ja Sams, M. (2003). *Ajatus ohjaa tietokonetta*. Tiede 2/2003, 46-47.
- Maybury, M. T. (toim.) (1997). *Intelligent Multimedia Information Retrieval*. AAAI Press
- Minsky, M. (2000). *Commonsense-Based Interfaces*. Communications of the ACM August 2000 (Vol.43, No.8).
- Nass, C. ja Gong, L. (2000). *Speech Interfaces from an Evolutionary Perspective*. Communications of the ACM September 2000 (Vol.43, No.9).
- Neisser, U. (1981). *Kognitio ja todellisuus*. Weilin+Göös. (englannink. alkuteos 1980).
- Neuss, R. (2002). *Usability Engineering as Approach to Multimodal Human Machine Interaction*. Proceedings of the ISCA Tutorial and Research Workshop on Multi-Modal Dialogue in Mobile Environments, June 17-19, 2002, Germany.
- Nigay, L. ja Coutaz, J. (1993). *A design space for multimodal systems - concurrent processing and data fusion*. In INTERCHI '93 - Conference on Human Factors in Computing Systems, Amsterdam, 172-178. Addison Wesley, 1993.
- Nielsen, J. (1993). *Usability engineering*. Academic Press.
- Norman, D. ja Draper, S. W. (1986). *User centered system design*. Lawrence Erlbaum Associates.
- Norman, D. (1989). *Miten avata mahdottomia ovia?*. Weilin+Göös. (englannink. alkuteos 1988).
- Norman, D. (1993). *Things That Make Us Smart*. Addison-Wesley.

Norman, D. (1999). *Affordance, Conventions and Design*. Originally published as affordance, conventions, and design in the May 1999 issue of interactions, 38-43 [<http://www.jnd.org/dn.mss/affordances-interactions.html> 14.2.2002]

Oviatt, S. ja Olsen, E. (1994). *Integration Themes in Multimodal Human-Computer Interaction*. In Proceeding of the International Conference on Spoken Language Processing, volume 2, Acoustical Society of Japan, 551-554.

Oviatt, S. (1996). *Multimodal interfaces for dynamic interactive maps*. Proceedings of the Conference on Human Factors in Computing Systems (CHI'96), ACM Press, New York, 1996, 95-102.

Oviatt, S., DeAngeli, A., Kuhn, K. (1997). *Integration and synchronization of input modes during multimodal human-computer interaction*. Proceedings of the SIGCHI conference on Human factors in computing systems, March 22-27, 1997, Atlanta, 415-422.

Oviatt, S. (1999). *Ten myths of multimodal interaction*. Communications of the ACM November 1999.

Oviatt, S. ja Cohen, P. (2000a). *Multimodal Interfaces That Process What Comes Naturally*. Communications of the ACM March 2000 (Vol.43, No.3).

Oviatt, S. (2000b). *Taming recognition Errors with a Multimodal Interface*. Communications of the ACM September 2000 (Vol.43, No.9).

Oviatt, S., Cohen, P., Wu, L., Vergo, J., Duncan, L., Suhm, B., Bers, J., Holzman, T., Winograd, T., Landay, J., Larson, J. ja Ferro, D. (2000c). *Designing the user interface for multimodal speech and gesture applications: State-of-the-art systems and research directions*. Human Computer Interaction (vol. 15, no. 4), 263-322.

Oviatt, S. (2001). *Designing robust multimodal systems for universal access*. Proceedings of the 2001 EC/NSF workshop on Universal accessibility of ubiquitous computing: providing for the elderly, May 22-25, 2001, Alcácer do Sal, Portugal.

Picard, R. W. (2000). *Affective Perception*. Communications of the ACM March 2000 (Vol.43, No.3).

Raisamo, R. (1999). *Multimodal Human-Computer Interaction: A Constructive and Empirical Study*. Väitöskirja. Tampereen Yliopisto, Tietojenkäsittelytieteiden laitos. [Myös: <http://acta.uta.fi/pdf/951-44-4702-6.pdf>, 7.4.2003].

Schomaker, L., Nijtmans, J., Camurri, A., Lavagetto, F., Morasso, P., Benôit, C., Guiard-Marigny, T., Le Goff, B., Robert-Ribes, J., Adjoudani, A., Defée, I., Münch, S., Hartung, K. ja Blauert, J. (1995). *A Taxonomy of Multimodal Interaction in the Human Information Processing System*. A Report of the ESPRIT project 8579 MIAMI. [<http://hwr.nici.kun.nl/~miami/taxonomy/taxonomy.html>, 7.4.2003].

Shneiderman, B. (1998). *Designing the User Interface (kolmas painos)*. Addison-Wesley.

Shneiderman, B. (2000). *The Limits of Speech Recognition*. Communications of the ACM September 2000 (Vol.43, No.9).

Sinkkonen, I., Kuoppala, H., Parkkinen, J. ja Vastamäki R. (2002). *Käytettävyyden psykologia*. It Press/Edita.

Suhm, B., Myers, B. ja Waibel, A. (2001). *Multimodal error correction for speech user interfaces*. ACM Transactions on Computer-Human Interaction March 2001 (Vol. 8, No. 1).

Suchman, L. (1987). *Plans and situated actions*. Cambridge University Press.

Suchman, L. (2000). *Human/Machine reconsidered*. Lancaster University. [<http://www.comp.lancs.ac.uk/sociology/soc040ls.html>, 15.5.2004].

Sutcliffe, A. (2000). *Designing Multimedia Presentations*. CHI 2000 Tutorial.

Talouselämä (2003). *Kaikki aistit käyttöön*. Talouselämä 11/2003, 24.

Tietotekniikan liitto ry. *ATK-sanakirja*. [<http://www.tt-tori.fi/atk-sanakirja/index.htm>, 15.4.2002].

Tuuri, K. (2002). *MMA A26 MULTIMEDIATEKNIIKAN PERUSTEET, johdantoluennon ka/vot*. Jyväskylän yliopisto. [Myös: <http://www.multimedia.jyu.fi/opinnot/kurssimateriaalit/mmaa26/johdanto/index.htm>, 15.4.2002].

Vaittinen, T. (2003). *Guideline-supported user-centered design of multimodal speech-enabled TV-guide*. Diplomityö. Helsingin teknillinen korkeakoulu, Tietotekniikan osasto.

Weinschenk S. ja Barker D. T. (2000). *Designing Effective Speech Interfaces*. John Wiley & Sons.